DOCUMENT RESUME

ED 057 086 TM 000 929

AUTHOR Thompson, Raymond E.
TITLE Investigations of the Appropriateness of the College Board Science Achievement Tests for Students of Different High School Science Courses.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO TDR-71-2
PUB DATE Sep 71
NOTE 79p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Achievement Tests; Biology; Chemistry; Course Evaluation; *Educational Background; *High School Students; *Performance Factors; Physics; Rating Scales; *Science Courses; Science Tests; Statistical Analysis; Teacher Attitudes; Test Bias; Test Interpretation; Test Results; Test Reviews
IDENTIFIERS *College Board Science Achievement Tests

ABSTRACT

Results of teacher ratings of test questions, performance of students, and analyses of achievement test scores indicate that the College Board Science Achievement Tests are now equally appropriate for students in both regular and special courses in biology, chemistry, and physics. (MS)

ED057086

COLLEGE ENTRANCE EXAMINATION BOARD
RESEARCH AND DEVELOPMENT REPORTS

RDR-71-72 No. 2

•

TEST DEVELOPMENT REPORT
TDR-71-2 SEPTEMBER 1971

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Investigations of the Appropriateness of the College Board Science Achievement Tests for Students of Different High School Science Courses

•

Raymond E. Thompson
Science Department
Test Development Division, ETS

TM 000 929

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY Educational Testing Service
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

COPYRIGHT © 1971 BY EDUCATIONAL TESTING SERVICE. ALL RIGHTS RESERVED.

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY
BERKELEY, CALIFORNIA

ED057086

COLLEGE ENTRANCE EXAMINATION BOARD

Research and Development Reports

RDR-71-72, No. 2

INVESTIGATIONS OF THE APPROPRIATENESS OF THE COLLEGE BOARD SCIENCE ACHIEVEMENT TESTS FOR STUDENTS OF DIFFERENT HIGH SCHOOL SCIENCE COURSES

Raymond E. Thompson, Science Department, Test Development Division, ETS

Test Development Report

TDR-71-2, September 1971

Educational Testing Service

Princeton, New Jersey

Berkeley, California

TABLE OF CONTENTS

	Page
Acknowledgments	i
I. Summary	1
II. Context	3
A. Background	3
B. Approach	4
C. Implications for Achievement Testing and Course Development	4
III. Purpose	6
IV. Design	6
A. Identification of Examinees in Different Course Categories	6
B. Checks on the Accuracy of Examinee Responses Regarding the Courses They Studied	9
C. Ratings of the Appropriateness of Test Questions for Students of Different Courses by Teachers of the Courses	10
D. Actual Test Performance of Examinees in Different Courses	11
E. Adjusted Test Performance of Examinees in Different Courses	11
V. Data and Discussion	14
A. Numbers and Characteristics of Examinees in Different Course Categories	15
B. Accuracy of Examinee Responses Regarding Courses Studied	21
C. Ratings of the Appropriateness of Test Questions for Students of Different Courses by Teachers of the Courses	25
D. Actual Performance of Examinees in Different Courses on:	29
1. Complete Science Achievement Tests	
2. Subtests of the Achievement Tests Rated Appropriate for Different Courses	
3. Scholastic Aptitude Test (SAT)	
E. Adjusted Performance of Examinees in Different Courses on the Science Achievement Tests after taking Account of Performance on Subtests of Achievement Tests Rated Appropriate for Different Courses and SAT	52
VI. Conclusions	62
List of Figures	66
List of Tables	67
Appendixes	69
I. Science Achievement Tests Question Rating Form.	69
II. Directions for Rating Test Questions.	70
III. Total Numbers of Examinees Taking Each Science Achievement Test per School Year	71
IV. Numbers and Percentages of Examinees in Different Courses.	72
V. Mean Test Scores of Students in Different Course Categories.	74
Bibliography.	75

ACKNOWLEDGMENTS

The contributions of William H. Angoff, Henry S. Dyer, Frank J. Fornoff, Lela Franklin, Sushil Kumar, Richard S. Levine, Donald J. Malcolm, Gary L. Marco, Charles T. Myers, William B. Schrader, E. Elizabeth Stewart, David W. Stickell, Frances Swineford, and Richard W. Watkins to these investigations are gratefully acknowledged.

Investigations of the Appropriateness of the College Board Science Achievement Tests for Students of Different High School Science Courses

I. SUMMARY

The appropriateness of the College Board Achievement Tests in Biology, Chemistry, and Physics for students of different high school science courses was investigated. These investigations involved six Biology, twelve Chemistry, and eight Physics Achievement Test Forms introduced in the 1960's. The comparative fairness of the tests for various pairs of courses was analyzed. These pairs were: BSCS[1]-Blue and Regular Biology, BSCS-Green and Regular Biology, and BSCS-Yellow and Regular Biology; CBA[2] and Regular Chemistry and CHEMS[3] and Regular Chemistry; and PSSC[4] and Regular Physics.

For most test forms three indications of their appropriateness for the courses in each relevant pair were secured. The first was ratings of test questions by teachers of the two courses. The second indication came from the actual performance of students of the two courses on the complete science achievement test, on questions in the tests rated appropriate for both courses by teachers, on the College Board Scholastic Aptitude Test - Verbal Section, and on SAT - Mathematical Section. The third indication came from an analysis of the achievement test scores of students of the two courses after adjusting for their differing performances on the concomitant measures of the appropriate-for-both questions, SAT-V, and SAT-M.

In interpreting the three indications of appropriateness, the following standards were taken to show bias: (1) mean teacher ratings differing significantly at the .05 level; (2) actual means on the achievement tests differing in one direction with actual means on the concomitant measures differing in the opposite direction; and (3) adjusted achievement test means differing by more than 15 scale score points after controlling for performance on the concomitant measures through the analysis of covariance.

---

[1]Biological Sciences Curriculum Study

[2]Chemical Bond Approach Project

[3]Chemical Education Material Study

[4]Physical Science Study Committee

In biology, teacher ratings indicated bias in favor of Regular Biology in the first two test forms studied. These indications of bias were not borne out by the test performance of students, however. In fact, in biology the lone indication of bias revealed by analysis of student performance showed that one of these first two test forms was biased in favor of BSCS-Yellow. On the last four test forms investigated, there was only one indication of bias; that came from teacher ratings and indicated that a test form introduced in 1966 was biased in favor of BSCS-Yellow.

In chemistry, there were indications of bias in seven of the twelve test forms as follows:

Form introduced in:	Biased in favor of:	Evidence from:
1961	regular over CHEMS	actual performance
1962	regular over CBA	actual performance
1963	regular over CBA	teacher ratings, actual and adjusted performance
	regular over CHEMS	teacher ratings, actual and adjusted performance
1964	regular over CBA	teacher ratings
	regular over CHEMS	teacher ratings
1964	regular over CBA	teacher ratings and adjusted performance
	regular over CHEMS	teacher ratings and adjusted performance
1966	regular over CHEMS	actual performance
1966	CHEMS over regular	teacher ratings

Teacher ratings on each course were not secured for forms prior to 1963; performance data are not available for the first 1964 form listed above. For the last five test forms studied, only the two indications of bias on 1966 forms noted above were revealed.

In physics, there were indications from teacher ratings of bias in favor of Regular Physics on two test forms, one introduced in 1965 and one in 1966. Actual performance data supported the ratings for the 1965 form, but not the 1966 form. Adjusted performance data indicated no bias in any of the eight forms investigated.

The evidence supports the general conclusion that the tests are now equally appropriate for students of regular and special courses in biology, chemistry, and physics.

II. CONTEXT

The problem investigated was: Do students who are equivalent in scholastic aptitude and science ability but who have studied different high school science courses in a given subject earn equivalent scores on the College Board Science Achievement Test in that subject?

A. Background

Beginning in the late 1950's and continuing to the present several new high school science courses were developed by teams of scientists and science teachers. Since the new courses were thought by many to represent departures from existing courses in terms of content, approach, and emphasis, there was and is interest in and concern about the appropriateness of the College Board Science Achievement Tests for students of these new courses.

The first science course to attract substantial numbers of students who later took a College Board Science Achievement Test was the physics course developed by the Physical Science Study Committee (PSSC). This course was soon followed by the two chemistry courses developed by the Chemical Bond Approach Project (CBA) and the Chemical Education Material Study (CHEMS); and the three biology courses known as the Blue, Green, and Yellow Versions developed by the Biological Sciences Curriculum Study (BSCS).

Fornoff (1962) noted that the performances of the first PSSC students who took the Physics Achievement Test, in March 1958, showed that the test did not adequately measure their achievement. The Board's Committee on Examinations, therefore, authorized special physics tests for PSSC students. Such special tests were offered in March of 1959, 1960, 1961, 1963, and in December of 1961. Fornoff pointed out that the existence of two physics tests solved the problem of not providing a test closely matched to course objectives, but it introduced two others. The first problem introduced was that equal scores on the two tests did not necessarily represent equal achievement in physics; the second problem was that some students took the wrong test and such mistakes cannot readily be detected and corrected.

B. Approach

In order to avoid these problems, special tests were not developed for students of any of the courses that followed PSSC. Instead, the approach has been to develop single tests in each subject that would be appropriate for students of different courses. Principles that guide the committees of examiners in the development of such tests include the following: Most of the questions in a test deal with topics given major emphasis in most courses. The measurement of abilities that should be developed in all science courses is emphasized. Questions on topics more likely to be taught in one course are balanced by other questions on topics more likely to be treated in other courses. Questions on topics that may be unfamiliar to students of some courses are presented with background information so that a good science student should be able to answer them even though he has not studied the topics in detail. In many instances, the committees of examiners have data on the difficulty and discriminating power of proposed questions for students of different courses secured through the pretesting of questions.

C. Implications for Achievement Testing and Course Development

Coffman (1971) notes that colleges may require applicants to submit scores on certain College Board Achievement Tests for one or more of the following purposes: (1) to aid in certifying that a candidate has or has not achieved a level of competence in a subject; (2) to assist in placing students in a college course sequence; and (3) to make predictions of college performance in combination with other information. These purposes are probably best served by a single test in each of the three major high school science subjects for the following reasons. Most colleges are concerned about student competency in a science subject; e.g., biology, rather than in a specific biology course; e.g., BSCS (Yellow Version). Most colleges do not have different sequences of courses for students of different high school courses in that subject. Finally, there is no reason to believe that special achievement tests designed for particular courses would make any greater contribution to the prediction of general college performance than single achievement tests designed for students of all courses in that subject.

One may ask why students of a particular course should not earn higher scores on a test than students of another course. After all, is not the purpose of developing new courses the improvement of the science achievement of students? Tests could be developed that focused on the unique content of a particular course. No doubt students of that particular course would be at an advantage on such tests. But the College Board Science Achievement Tests cannot be so developed, if they are to serve their purposes. They must focus primarily on that broad range of content that is common to all widely used courses. One price that must be paid in order to realize those purposes is that the unique content of each avant-garde course cannot be immediately included in the tests. But this seems a small and legitimate price to pay to meet the purposes of these tests.

For an outline of the issues discussed in this section see Angoff (1971).

III. PURPOSE

The purpose of these investigations was to evaluate the appropriateness of the College Board Science Achievement Tests for students of different high school science courses. The purpose was not to find out if one course was better than another. Instead, the purpose was to determine if students of a given course were at a disadvantage on the relevant science achievement test.

IV. DESIGN

Six Biology, twelve Chemistry, and eight Physics Achievement Test Forms were studied. The comparative appropriateness of each test form for relevant pairs of courses was investigated. The pairs of courses were: BSCS-Blue and Regular Biology, BSCS-Green and regular, and BSCS-Yellow and regular; CBA and Regular Chemistry, CHEMS and regular; and PSSC and Regular Physics.

For most test forms three measures of appropriateness were obtained. These measures were: (1) ratings of achievement test questions by teachers of both courses; (2) actual performance of students of both courses on the achievement test and on concomitant variables; and (3) adjusted achievement test scores resulting from controlling for performance on the concomitant variables. The concomitant variables were subsets of the achievement tests rated highly and equally appropriate for both courses by the teachers, and the Scholastic Aptitude Test-Verbal and Mathematical Sections.

The following indications of bias were adopted: (1) mean teacher ratings differing significantly at the .05 level; (2) actual means on the achievement tests differing in one direction with actual means on the concomitant variables differing in the opposite direction; and (3) adjusted achievement test means differing by more than 15 scale score points after controlling for performance on the concomitant variables through the analysis of covariance.

Dyer, Levine, and Watkins were active in the initial design of these investigations. Malcolm (1962) was responsible for the first and Stickell (1965) for the second investigation of the Biology Test forms. All of the investigations have followed essentially the same pattern. The steps in an investigation are outlined below.

A. Identification of Examinees in Different Course Categories

To determine the science course a student was studying or had studied when he took a science achievement test, the following communications were printed on the front covers of the Biology, Chemistry, and Physics Tests, respectively.

Biology Test

The question below pertains to whether or not you have taken some special courses in biology which are offered by some schools. You are to indicate your answer by blackening ONE and ONLY ONE space in the group of nine spaces labeled Q on your answer sheet. Read the question and the statements under it and blacken the space that corresponds to the statement that applies to you. Your answer to this question will be used for research purposes only and will not influence your score on the test.

Question: Have you taken, or are you now taking, one of the courses in biology developed by the Biological Sciences Curriculum Study (BSCS)?

NOTE: If you were in a BSCS course, you either used paper-covered textbooks which had the symbol shown at the right imprinted on the covers or you used one of the hard-covered textbooks whose titles are given below. If you used the paper-covered books, the color of the covers indicated the version of the course you took.

Statements

Space 1. Yes, I took the Green Version of the BSCS course. (Title of hard-covered edition: High School Biology)

Space 2. Yes, I took the Yellow Version of the BSCS course. (Title of hard-covered edition: Biological Science: an Inquiry into Life)

Space 3. Yes, I took the Blue Version of the BSCS course. (Title of hard-covered edition: Biological Science: Molecules to Man)

Space 4. Yes, I took a BSCS course but I am not sure of the version.

Space 5. I am not sure whether or not I took a BSCS course.

Space 6. No. I have not taken a BSCS course.

Spaces 7-9. These spaces are to be left blank.

Chemistry Test

In the group of nine spaces labeled Q, you are to blacken ONE and ONLY ONE space, as described below, to indicate how you obtained your knowledge of chemistry. The information that you provide will not influence your score on the test.

Space 1. I am now taking, or have taken, the chemistry course known as the Chemical Bond Approach Course (CBA). (If this applies to you, the symbol shown at the right will be familiar to you and you will have used either paper-covered textbooks with pages the same size as those in this test booklet or a hard-covered textbook titled Chemical Systems.)

Space 2. I am now taking, or have taken, the chemistry course known as the Chemical Education Material Study Course (CHEM Study). (If this applies to you, the symbol shown at the right will be familiar to you and you will have used either paper-covered textbooks with pages the same size as those in this test booklet or a hard-covered textbook titled Chemistry: An Experimental Science.)

Space 3. I am not sure if I am taking, or have taken, either the CBA or the CHEM Study Course.

Space 4. I am not taking, or have not taken, either the CBA or the CHEM Study Course.

Spaces 5-9. These spaces are to be left blank.

Physics Test

To provide information on your training in physics, please read the following statements and select the one statement that applies to you. Then, in the group of nine spaces labeled Q on your answer sheet, blacken the one space whose number corresponds to that of the statement you selected. The information that you are asked to supply is for research purposes only and will have no effect on your test score.

Space 1:

I am now taking, or have taken, the physics course prepared by the Physical Science Study Committee (PSSC). (If you took this course, you used a textbook with a stroboscopic photograph of a bouncing ball on its cover.)

Space 2:

I am not sure if I am taking, or have taken, the PSSC course.

Space 3:

I am not taking, or have not taken, the PSSC course.

Spaces 4-9:

These spaces are to be left blank.

The following chart indicates which courses were included in these studies.

Subject	Courses	Space Marked by Student on Test Cover
Biology	BSCS, Blue Version	3
	BSCS, Green Version	1
	BSCS, Yellow Version	2
	Regular Biology	6
Chemistry	CBA	1
	CHEM Study	2
	Regular Chemistry	4
Physics	PSSC	1
	Regular Physics	3

Examinees were categorized as students of the regular course if they indicated that they were not students of the "special" courses.

B. Checks on the Accuracy of Examinee Responses Regarding the Courses They Studied

There has been concern about the accuracy of the examinee responses regarding courses studied. Several checks on the accuracy of these responses have been made. These checks have always involved asking staff members of schools attended by examinees to provide information on the science courses the examinees have studied. In most of these checks a list of examinees was sent to each school along with the courses the examinees said they studied. The schools were asked to verify or reject the examinees' claims.

One check focused on examinees who said they were not sure about their course, as well as those who made uninterpretable responses not in keeping with the directions, and those who failed to respond.

Since most of the comparisons of performance were based on groups of examinees identified solely on the basis of examinee responses, there were understandably some qualms about the validity of these comparisons. Hence, for the January and May tests of 1967 in all three subjects complete analyses of performance for samples based on student responses and samples based on school verifications were carried out.

C. Ratings of the Appropriateness of Test Questions for Students of Different Courses by Teachers of the Courses

These ratings of appropriateness by the teachers actually served two purposes: (1) identification of subsets of items in each science achievement test that could serve as an unbiased measure of achievement in that science; and (2) direct judgments of the appropriateness of tests for different courses.

The following procedure was used in securing the ratings and identifying the questions appropriate for both courses under comparison in all investigations except the early ones in chemistry. About ten teachers of course X and ten teachers of course Y at a rating conference were directed to answer all of the questions in the test (to insure careful consideration of questions) and to record their ratings of the appropriateness of each question for the course they represented. A teacher was asked to give each question one of the following three ratings: appropriate and emphasized, appropriate but not emphasized, or inappropriate. The numbers 2, 1, or 0, respectively, were associated with the ratings.

Appendix II contains a copy of the written directions for rating questions that was given to the teachers.

In order to serve as a rater of the appropriateness of test questions for course X, a teacher had to be actively engaged in teaching course X. Ordinarily a rater was a teacher of only course X. A few raters were active teachers of courses X and Y; but such teachers provided ratings for only one of those two courses.

The questions selected for the unbiased measure of achievement were those rated uniformly and equally high by both X and Y teachers. The selected questions were those that had ratings with high means and low variances for course X or course Y teachers considered separately, and equal mean ratings for both courses. At least 20 questions were selected for most test forms studied.

If the appropriateness of a test for courses X, Y, and Z was under study, one subset of questions appropriate for both X and Y was used in comparing the appropriateness of the full test for X and Y students. Another subset of questions appropriate for X and Z was used in comparing the appropriateness of the complete test for X and Z students. Usually there was considerable overlap between the subsets.

For the early investigations in chemistry, appropriateness ratings were obtained from CBA and CHEMS teachers only. It was assumed that the questions were appropriate for regular students and hence no ratings from teachers of regular courses were secured. In the early chemistry investigations, the raters were instructed to rate each question as either appropriate or inappropriate for the course they represented. Each rating of appropriate was assigned a value of 1 and each rating of inappropriate was assigned a value of 0. Then substantial numbers of questions that had relatively high appropriateness ratings were selected for the unbiased measure of achievement in chemistry.

D. Actual Test Performance of Examinees in Different Courses

For most of the comparisons, the scores of examinees in two courses on (1) a complete science achievement test, (2) a subset of the achievement test rated equally appropriate for both courses, and (3) the Scholastic Aptitude Test were used. Several of the comparisons in chemistry took no account of SAT performance. All of the comparisons except for the first two in chemistry were for science achievement tests given in January or May.

E. Adjusted Test Performance of Examinees in Different Courses

The final goal was to obtain the mean test scores on a given science achievement test form to be expected of two groups of examinees from two different courses who were equivalent in science ability and in scholastic aptitude. The statistical technique for obtaining these mean scores was the analysis of covariance.

According to Lindquist (1953) there are seven conditions to be met in the analysis of covariance before valid tests of the significance of differences between adjusted mean scores can be made. These seven conditions are:

1. Students in both course samples are drawn at random from the same parent population or selected from the same parent population only on the basis of scores on the concomitant variables; i.e., the appropriate-for-both questions and SAT.
2. Scores of students on the appropriate-for-both questions and SAT are not differentially affected by the courses the students studied.

3. The achievement test scores for both course samples are random samples of those for corresponding course populations.
4. The regression of achievement test scores on appropriate-for-both scores and SAT scores is the same for both course populations.
5. This regression is linear.
6. The distributions of achievement test scores for each of the two course populations are normal.
7. Both of these distributions have the same variance.

A discussion of how well the experimental design met each of these conditions follows:

1. With respect to the first condition, it could be assumed that a parent population consists of all students who take a given science achievement test form. The students in the regular and special course samples for this test were not drawn at random from this parent population nor were they selected from it only on the basis of scores on the appropriate-for-both questions and SAT. Students ended up in the regular or special course populations for the test on the basis of their indications of the courses they studied.
2. The second condition is that performances on the appropriate-for-both questions and on SAT are not differentially affected by the courses students studied. These effects are probably slight. If questions are judged by teachers of both courses to be appropriate for both courses, achievement on these questions is probably not differentially affected to an appreciable degree by the courses studied. Differential effects of courses on SAT performance are likely to be even smaller.
3. The third condition is that the achievement test scores for both samples in a comparison are random samples of those for the corresponding populations. This condition is contained in 1. If condition 1 is met, then condition 3 is surely met. Meeting of condition 3 is necessary but not sufficient for meeting condition 1. The regular and special course samples

were drawn systematically from regular and special course populations, respectively. Tests were made of the significance of differences between population and corresponding sample means. Even more important for a given comparison between course X and course Y was the following consideration: The difference between the X sample and population means should not differ significantly from the difference between the Y sample and population means. Figure 1 illustrates the criticality of this consideration.

FIGURE 1

Examples of Various Relationships Between Course X Sample and Population Means and Course Y Sample and Population Mean

The comparisons based on samples in examples 1 and 2 of Figure 1 will be valid. In example 3 the difference between the X sample and population means is markedly different from the difference between the Y sample and population means. Here the comparison based on samples will be invalid. Comparisons in which the difference between the regular sample and population means was significantly different from the difference between the special sample and population means at the .01 level were ruled out of consideration.

4. Two different tests of condition 4 were made for each comparison. All comparisons in which errors of estimate or slopes of the regression lines for the two samples differed significantly at the .01 level were ruled out of consideration.

5 and 6. Conditions 5 and 6 will be assumed.

7. Condition 7 is contained in condition 4. The satisfaction of condition 7 was tested for by the tests of the significance of the difference between the errors of estimate about the regression lines for the two course samples under comparison.

Another condition to be met is that the correlations between achievement test scores and scores on the concomitant variables be positive and substantial. It is reasonable to adjust achievement test scores only for performance on measures that are closely related to it. Correlations between achievement test scores and scores on the concomitant variables were found for both samples in each comparison.

V. DATA AND DISCUSSION

The data compiled will be presented and discussed in the following order:

A. Numbers and Characteristics of Examinees in Different Courses

B. Accuracy of Examinee Responses Regarding Courses Studied

C. Ratings of the Appropriateness of Test Questions for Students of Different Courses by Teachers of the Courses

D. Actual Performances of Examinees in Different Courses on:

1. Complete Science Achievement Tests
2. Subtests of the Achievement Tests Rated Appropriate for Different Courses
3. Scholastic Aptitude Test (SAT)

E. Adjusted Performances of Examinees in Different Courses on the Science Achievement Tests After Taking Account of Performances on Subtests of Achievement Tests Rated Appropriate for Different Courses and SAT

A. Numbers and Characteristics of Examinees in Different Course Categories

The total numbers of examinees taking the Biology, Chemistry, and Physics Achievement Tests in recent school years are given in Appendix III.

The numbers of examinees studying the different courses in each science are given in Appendix IV. These data and some additional data all drawn from Swineford's statistical reports (1962-1969) are depicted in Figures 2-4. In these three graphs the uncertains included those examinees who indicated that they were not sure about the course they took as well as those who failed to respond and a few who marked spaces that were to be left blank. Those biology examinees who indicated they studied a BSCS course but were uncertain of the version were included in the BSCS (All Versions) category.

Figure 2

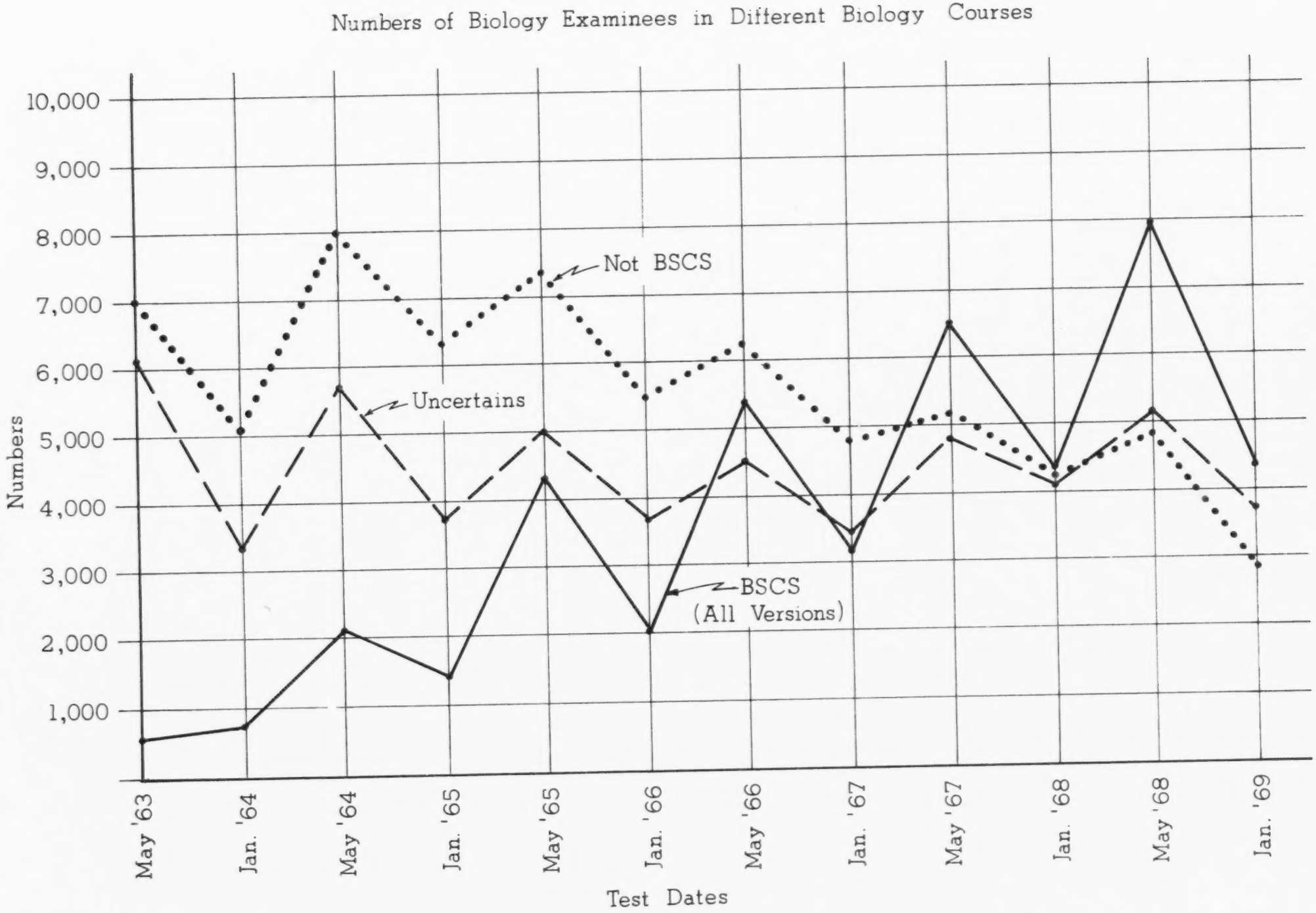

Numbers of Biology Examinees in Different Biology Courses
10,000
9,000
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
Numbers
Not BSCS
Uncertains
BSCS
(All Versions)
May '63
Jan. '64
May '64
Jan. '65
May '65
Jan. '66
May '66
Jan. '67
May '67
Jan. '68
May '68
Jan. '69
Test Dates

Figure 3

Numbers of Chemistry Examinees in Different Chemistry Courses

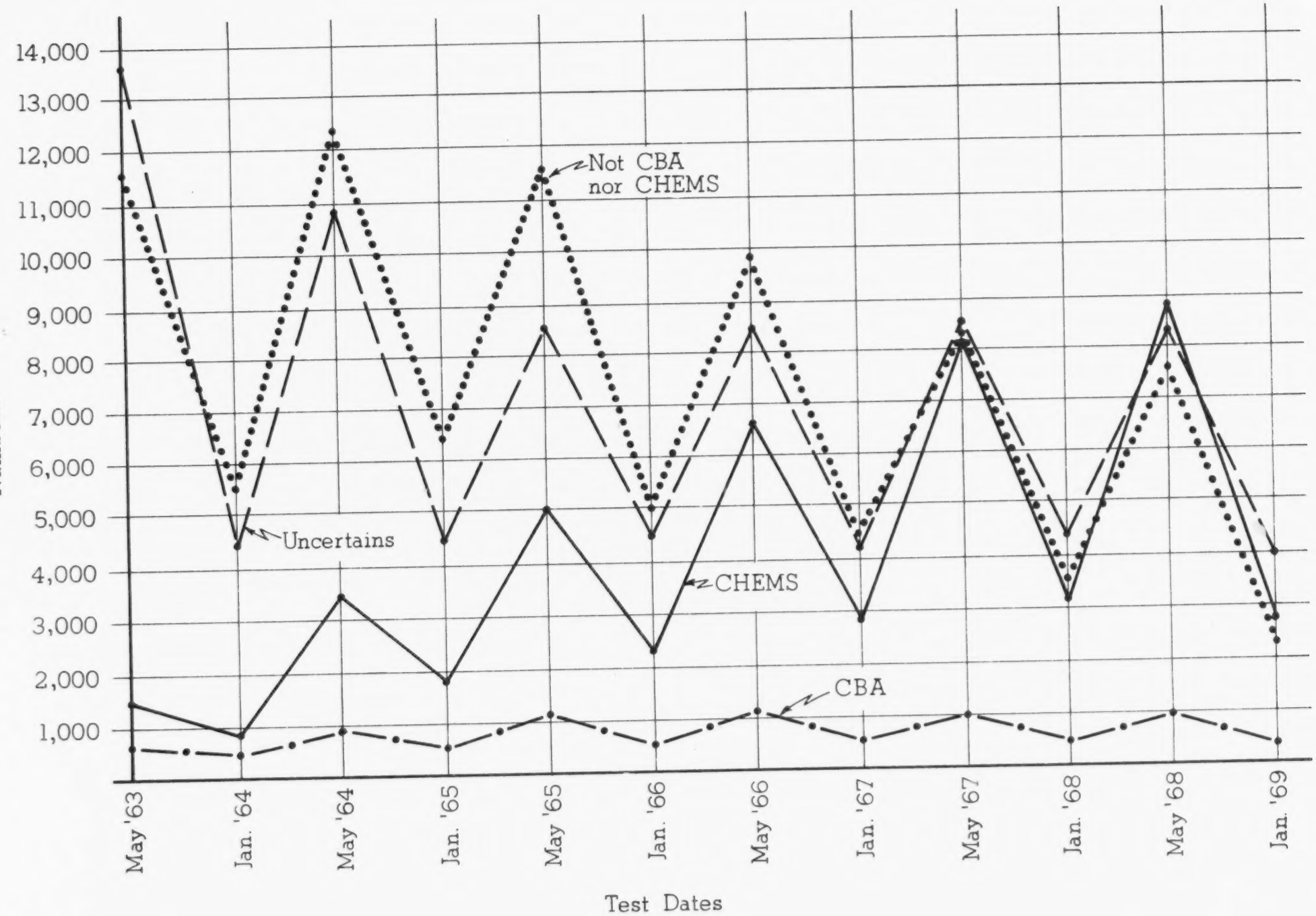
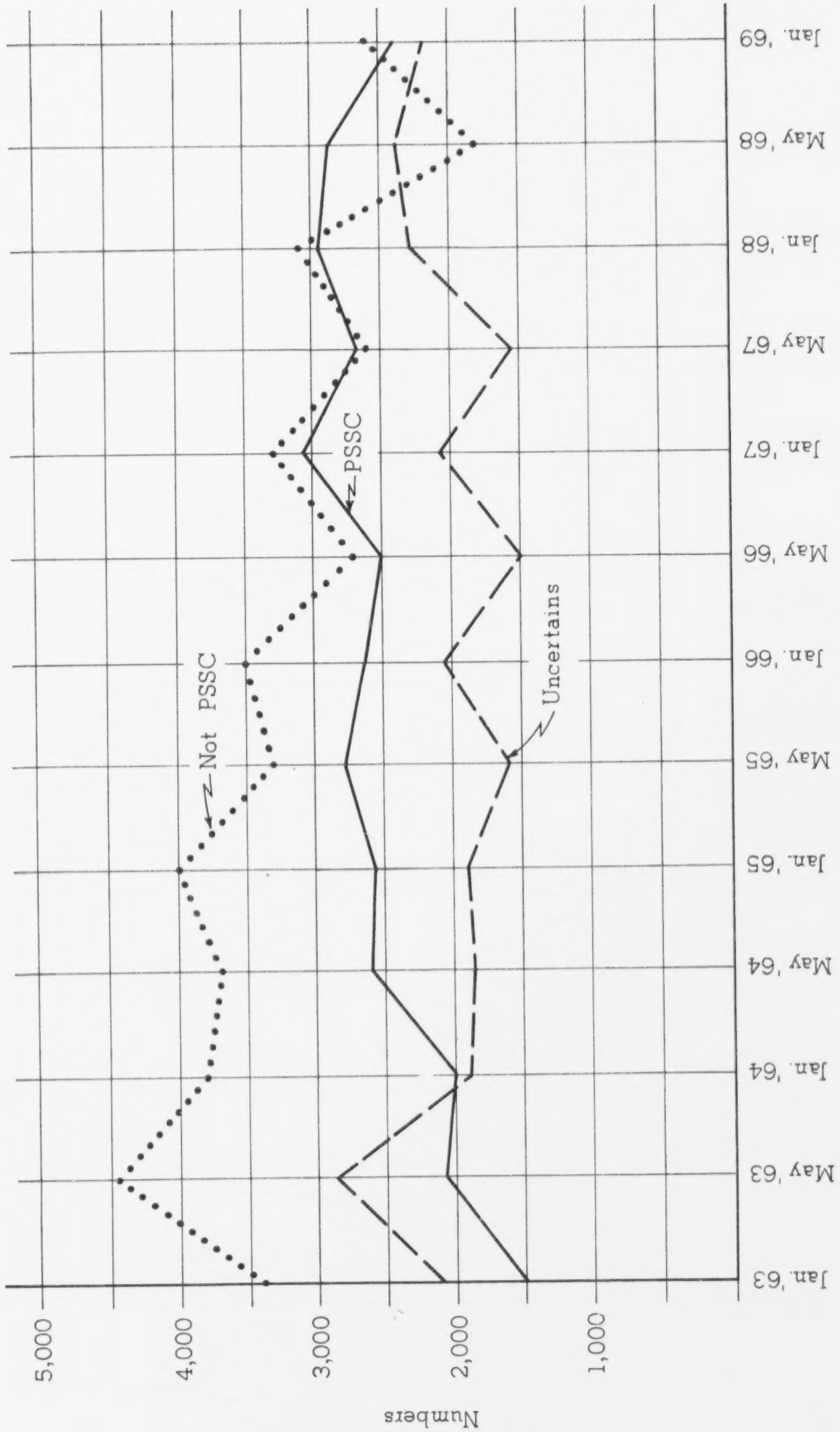


Figure 4

Numbers of Physics Examinees in Different Physics Courses



Figures 2-4 reveal that more students take the Biology and Chemistry Tests in May than in January; and that for biology, chemistry, and physics, students in special courses constitute a larger proportion of the examinee group in May than in January. Data presented by Fornoff, Kastrinos, and Thompson (all 1969) show that compared to the January examinees the May examinees in all three subjects include more juniors, more students from independent schools, more residents of the Northeast, and more students of the new courses.

One generalization supported by the data is that the numbers of examinees who have studied the leading new courses in each subject has been increasing, whereas the number who have studied the regular courses has been decreasing.

The regular course category is a catch-all for all examinees who did not think they took one of the new courses that is specifically named. The regular courses are not as diverse as one might think, however. Drawing on data reported by Fornoff, Kastrinos, and Thompson (all 1969) the following Table 1 on textbooks used in regular courses by samples of examinees in 1965-1966 was constructed. It clearly shows that for both mid-school-year (1965-1966) senior examinees and end-of-school-year (1965-1966) junior examinees in biology, chemistry, and physics the Holt, Rinehart and Winston series of textbooks titled Modern Biology, Modern Chemistry, and Modern Physics, respectively, were the predominant texts in regular courses.

TABLE 1

Textbooks Used in Regular Courses by Samples of Mid-School-Year (1965-66) Senior Examinees and End-of-School-Year (1965-66) Junior Examinees in Biology, Chemistry, and Physics

Subject	Leading Regular-Course Textbooks	Mid-Year Senior Examinees		End-of-Year Junior Examinees	
		Total No. Reporting	Percentage Using Each Reg. Text	Total No. Reporting	Percentage Using Each Reg. Text
Biology		1,554		598	
	1. Modern Biology		43%		36%
	2. Exploring Biology, Science of Living Things		5		6
Chemistry		1,331		678	
	1. Modern Chemistry		56		44
	2. Chemistry and You		5		4
Physics		1,312		706	
	1. Modern Physics		42		30
	2. Physics, an Exact Science		4		2
	3. Exploring Physics		3		3

B. Accuracy of Examinee Responses Regarding Courses Studied

Several checks were made on the accuracy of the examinee responses to the questions about courses studied that appeared on the front covers of the tests.

The first data on this matter were compiled by Stickell (1965) for the BSCS biology students who took the May 1963 Biology Test and are shown in Table 2.

TABLE 2

Accuracy of Examinee Responses Regarding BSCS Biology Courses Studied
May 1963 Biology Test

Course	Number of Examinees Who Said They Studied Course	Number Whose Responses Were Verified by Their Schools	Number Whose Responses Were Denied by Their Schools	Number Whose Responses Were Neither Verified Nor Denied	Percentage Agreement*
BSCS Blue	250	226	11	13	95%
BSCS Green	95	55	28	12	66
BSCS Yellow	170	123	43	4	74

*The percentage agreement was found by dividing the number whose responses were verified by the number whose responses were either verified or denied and multiplying by 100.

The next data on the accuracy of student responses were compiled for the May 1963 and January 1964 Physics Tests; these data are shown in Table 3.

TABLE 3

Accuracy of Examinee Responses Regarding Physics Courses Studied: May 1963 and January 1964 Physics Test

Test Date	Course	Number of Examinees Who Said They Studied Course	Number of Examinees Whose Schools: Were Queried About Examinee Response	Replied to Query	Verified Examinee Response	Denied Examinee Response	Neither Verified nor Denied Examinee Response	Percentage Agreement*
May 1963	PSSC	2,121	368	346	333	11	2	97%
	Not PSSC	4,456	355	266	260	0	6	100
January 1964	PSSC	2,036	320	259	240	19	0	93
	Not PSSC	3,825	318	240	229	1	10	100

*The percentage agreement was found by dividing the number whose responses were verified by the number whose responses were either verified or denied and multiplying by 100.

The most extensive check on the accuracy of student responses regarding courses studied was made for the Biology, Chemistry, and Physics Tests given in January and in May of 1967. The data from this check are presented in Table 4.

TABLE 4

Accuracy of Examinee Responses Regarding Science Courses Studied: January and May 1967 Biology, Chemistry, and Physics Tests

Test Date	Course	Number in a Sample of Examinees who Said They Studied Course	Number Whose Responses Were Verified by Their Schools	Number Whose Responses Were Denied by Their Schools	Number Whose Responses Were Neither Verified Nor Denied	Percentage Agreement*
Jan. 1967	BSCS-Blue	500	378	63	59	86%
	BSCS-Green	500	343	91	66	79
	BSCS-Yellow	496	370	50	76	88
	Regular Biology	498	271	138	89	66
	CBA	439	260	115	64	69
	CHEMS	500	371	64	65	85
	Regular Chemistry	500	404	15	81	96
	PSSC	500	404	40	56	91
	Regular Physics	498	392	31	75	93
May 1967	BSCS-Blue	500	420	35	45	92
	BSCS-Green	500	375	76	49	83
	BSCS-Yellow	500	432	33	35	93
	Regular Biology	499	314	87	98	78
	CBA	500	307	131	62	70
	CHEMS	500	403	47	50	90
	Regular Chemistry	500	415	26	59	94
	PSSC	500	392	41	67	91
	Regular Physics	500	392	28	80	93

*The percentage agreement was found by dividing the number whose responses were verified by the number whose responses were either verified or denied and multiplying by 100.

The percentage of examinees in a course category who apparently do not belong there is fairly high in some cases; e.g., 34% in Regular Biology on the January 1967 Biology Test, and 31% in CBA on the January 1967 Chemistry Test. The presence of these misplaced examinees in the samples could adversely affect the validity of comparisons of the performance of the groups on the tests. Hence, the analyses of the 1967 tests in all three subjects became of critical importance because for these tests complete analyses of the performance of student-response samples and of school-verified samples were carried out. If the results for the school-verified samples were similar to the results for the student-response samples, then it could be inferred that the analyses of the earlier tests, based on student-response samples only, were valid. The results for the 1967 school-verified and student-response samples are presented later in connection with the complete results for all the samples.

An investigation of the courses actually studied by examinees who indicated they were uncertain of their course was made for the May 1964 Physics Test. Of the 8,151 examinees for this test, 1,849 fell in the uncertain category. The uncertains included 1,165 who specified they were not sure what course they studied, 9 who made uninterpretable responses not in keeping with the directions, and 675 who failed to respond at all. A systematic sample numbering 380 was drawn from the 1,849. The schools of 351 of these 380 students were asked to indicate what course the students studied. The schools of the remaining 29 were not contacted because school addresses were not available for 26, two were in foreign countries, and one was a college student. Of the 351 students whose schools were contacted, school replies were received for 302, of which 24 (8%) were designated as students of the PSSC Physics Course and 278 (92%) were designated as students of some other physics course; i.e., the regular course.

C. Ratings of the Appropriateness of Test Questions for Students of Different Courses by Teachers of the Courses

Each question in each test form was rated by teachers of the relevant courses as appropriate and emphasized (2 points), appropriate but not emphasized (1 point), or inappropriate (0). An exception to this procedure in chemistry is described below.

For each question the mean and standard deviation of the ratings were found for each relevant course. For each test form the mean and standard deviation of the mean question ratings for each relevant course were determined. In Tables 5, 6, and 7 the number of teacher raters involved and the mean and standard deviation of their mean question ratings for the complete tests and for the highly and equally appropriate subsets are presented for biology, chemistry, and physics, respectively.

Six of the 18 comparisons in Table 5 show significantly different mean ratings at the .05 level. Five of these six significant differences in mean ratings appear on the first two tests and all five favor the regular course. The other significant difference appears on the fourth test and favors BSCS-Yellow. There are no significant differences in mean ratings on the last two tests. In making these tests of significance, the standard deviation of mean ratings was assumed to be .40 for the May 1963 test.

Appropriateness ratings for a dozen Chemistry Tests are presented in Table 6. For the first three tests, only ratings from CBA and CHEMS teachers were obtained; it was assumed that the tests were appropriate for the regular course. Questions in these first three tests were rated appropriate (1) or inappropriate (0). For the last nine tests, ratings on the 2, 1, 0 scale from CBA, CHEMS, and regular teachers are presented.

If one considers the last nine Chemistry Tests in Table 6; i.e., beginning with the May 1963 test, there are 18 paired course comparisons (two comparisons for each of nine tests). Seven of the 18 comparisons show significantly different mean ratings at the .05 level. Six of these seven significant differences in mean ratings appear on the first three tests and all six favor regular courses. Only one of the seven significant differences appears on the last six tests; that one comparison favors CHEMS. In

TABLE 5

Biology: Teacher Ratings of Appropriateness

Test Date	Courses Compared	No. of Teacher Raters	Mean Question Ratings: Complete Test — Number of Questions	Mean	Stand. Dev.	Highly and Equally Appropriate Subset of Test — Number of Questions	Mean	Stand. Dev.
May 1962	BSCS-Blue & Reg. Biol.	5 5	100	1.31* 1.55	.36 .40	19	1.90 1.93	.10 .09
	BSCS-Green & Reg. Biol.	5 5	100	1.42* 1.55	.46 .40	17	1.93 1.91	.10 .10
	BSCS-Yellow & Reg. Biol.	5 5	100	1.43* 1.55	.43 .40	20	1.93 1.91	.09 .10
May 1963	BSCS-Blue & Reg. Biol.	6 6	100	1.24* 1.57		22	1.91 1.87	.12 .12
	BSCS-Green & Reg. Biol.	6 6	100	1.51 1.57		17	1.92 1.95	.08 .08
	BSCS-Yellow & Reg. Biol.	3 6	100	1.39* 1.57		20	1.82 1.88	.16 .13
Jan. 1966	BSCS-Blue & Reg. Biol.	11 8	100	1.35 1.33	.51 .38	26	1.72 1.72	.24 .21
	BSCS-Green & Reg. Biol.	9 8	100	1.29 1.33	.48 .38	17	1.74 1.74	.25 .21
	BSCS-Yellow & Reg. Biol.	11 8	100	1.44 1.33	.38 .38	30	1.73 1.73	.18 .18
May 1966	BSCS-Blue & Reg. Biol.	11 8	100	1.31 1.22	.46 .52	22	1.73 1.73	.17 .20
	BSCS-Green & Reg. Biol.	9 8	100	1.31 1.22	.42 .52	21	1.74 1.74	.13 .21
	BSCS-Yellow & Reg. Biol.	11 8	100	1.50* 1.22	.42 .52	23	1.83 1.83	.11 .14
Jan. 1967	BSCS-Blue & Reg. Biol.	11 10	100	1.45 1.44	.41 .37	34	1.76 1.76	.20 .17
	BSCS-Green & Reg. Biol.	10 10	100	1.46 1.44	.32 .37	32	1.78 1.78	.14 .15
	BSCS-Yellow & Reg. Biol.	11 10	100	1.49 1.44	.29 .37	36	1.74 1.74	.15 .16
May 1967	BSCS-Blue & Reg. Biol.	11 10	100	1.55 1.61	.37 .28	32	1.81 1.81	.19 .12
	BSCS-Green & Reg. Biol.	10 10	100	1.56 1.61	.26 .28	39	1.76 1.76	.11 .17
	BSCS-Yellow & Reg. Biol.	11 10	100	1.61 1.61	.26 .28	50	1.79 1.79	.14 .15

*Means significantly different at .05 level

TABLE 6

Chemistry: Teacher Ratings of Appropriateness

Test Date	Courses Compared	No. of Teacher Raters	Mean Question Ratings: Complete Test — Number of Questions	Mean	Stand. Dev.	Mean Question Ratings: Highly and Equally Appropriate Subset of Test — Number of Questions	Mean	Stand. Dev.
Dec. 1961	CBA & Reg. Chem.	4	95	.77	.28	76	.89	.14
	CHEMS & Reg. Chem.	5	95	.72	.30	74	.86	.16
Mar. 1962	CBA & Reg. Chem.	6	100	.65	.35	67	.86	.18
	CHEMS & Reg. Chem.	5	100	.62	.36	67	.84	.17
May 1962	CBA & Reg. Chem.	5	95	.76	.29	71	.90	.13
	CHEMS & Reg. Chem.	6	95	.72	.27	70	.86	.15
May 1963	CBA & Reg. Chem.	9 10	95	1.30* 1.76*	.51 .29	20	1.82 1.92	.09 .11
	CHEMS & Reg. Chem.	10 10	95	1.40* 1.76*	.48 .29	20	1.83 1.92	.13 .11
Jan. 1964	CBA & Reg. Chem.	8 9	95	1.39* 1.55	.54 .38	28	1.84 1.82	.09 .14
	CHEMS & Reg. Chem.	9 9	95	1.40* 1.55	.57 .38	31	1.85 1.85	.15 .12
May 1964	CBA & Reg. Chem.	9 12	95	1.24* 1.56*	.53 .36	20	1.78 1.78	.10 .15
	CHEMS & Reg. Chem.	12 12	95	1.41* 1.56*	.49 .36	30	1.79 1.79	.14 .15
Jan. 1965	CBA & Reg. Chem.	9 9	95	1.43 1.56	.50 .44	20	1.88 1.88	.10 .10
	CHEMS & Reg. Chem.	10 9	95	1.54 1.56	.42 .44	20	1.92 1.88	.09 .10
May 1965	CBA & Reg. Chem.	7 7	95	1.58 1.55	.40 .43	20	1.89 1.90	.09 .11
	CHEMS & Reg. Chem.	8 7	95	1.54 1.55	.42 .43	20	1.91 1.90	.10 .11
Jan. 1966	CBA & Reg. Chem.	8 9	95	1.58 1.54	.40 .49	26	1.90 1.90	.05 .10
	CHEMS & Reg. Chem.	9 9	95	1.60 1.54	.38 .49	29	1.88 1.88	.12 .11
May 1966	CBA & Reg. Chem.	9 12	95	1.22 1.36	.51 .52	19	1.77 1.77	.10 .20
	CHEMS & Reg. Chem.	12 12	95	1.53* 1.36	.39 .52	31	1.82 1.82	.14 .14

*Means significantly different at .05 level

TABLE 6 (con't)

Test Date	Courses Compared	No. of Teacher Raters	Mean Question Ratings: Complete Test — Number of Questions	Mean	Stand. Dev.	Highly and Equally Appropriate Subset of Test — Number of Questions	Mean	Stand. Dev.
Jan. 1967	CBA &	9	90	1.60	.43	38	1.85	.11
	Reg. Chem.	11		1.55	.36		1.85	.14
	CHEMS &	11	90	1.65	.37	44	1.83	.16
	Reg. Chem.	11		1.55	.36		1.83	.15
May 1967	CBA &	9	90	1.58	.42	31	1.87	.11
	Reg. Chem.	11		1.62	.36		1.87	.11
	CHEMS &	11	90	1.67	.33	45	1.84	.17
	Reg. Chem.	11		1.62	.36		1.84	.13

chemistry there is convincing evidence from teacher ratings that the tests have evolved so as to be as appropriate for students of CBA and CHEMS courses as for students of regular courses.

Appropriateness ratings for eight Physics Tests are shown in Table 7. For the first two tests two different methods were used to pick the appropriate-for-both questions. In the first method the emphasis was on selecting questions that had PSSC and regular ratings with high means and low variances. In the second method the emphasis was on picking questions that had equal PSSC and regular ratings. The reader will recognize that these were two considerations in picking all sets of appropriate-for-both questions. For the first two Physics Tests the effect of emphasizing one or the other of these considerations was studied. The effect was appreciable as will be shown and discussed in Part E.

In Table 7 only two of the eight comparisons (January 1965 and May 1966) show significantly different mean ratings at the .05 level and both favor the regular course.

D. Actual Performances of Examinees in Different Courses on the:

1. Complete Science Achievement Tests
2. Subsets of the Achievement Tests Rated Appropriate for Different Courses
3. Scholastic Aptitude Test

The actual mean test scores of the examinees in different courses on the Science Achievement Tests are given in Appendix V. The mean scores in Appendix V are all on the same College Board Scale, which can extend from 200 to 800. Data on the actual test performance of samples of those examinees in the different course categories will be presented next. The mean scores of these samples will be set forth by subjects. Some of the samples for the early studies were not samples at all, but simply all the examinees for whom the relevant data could be assembled. For later studies systematic samples were constituted with answer sheets drawn at evenly spaced intervals from the answer sheets of all examinees in a given course. From many of these systematic samples the examinees for whom no SAT scores were available were excluded thus yielding reduced systematic samples. In other instances, reduced systematic samples were drawn at evenly spaced intervals from lists of

TABLE 7

Physics: Teacher Ratings of Appropriateness

Test Date	Courses Compared	No. of Teacher Raters	Mean Question Ratings: Complete Test — Number of Questions	Mean	Stand. Dev.	Mean Question Ratings: Highly and Equally Appropriate Subset of Test — Number of Questions	Mean	Stand. Dev.
						Emphasis on picking questions whose ratings had high means and low variances among the 10 raters for each course		
May 1963	PSSC &	10	75	1.40	.48	17	1.86	.19
	Reg. Phys.	10		1.51	.30		1.68	.19
Jan. 1964	PSSC &	10	75	1.50	.55	21	1.89	.20
	Reg. Phys.	10		1.52	.40		1.76	.13
						Emphasis on picking questions with equal ratings for both courses		
May 1963	PSSC &	10				19	1.59	.23
	Reg. Phys.	10					1.58	.28
Jan. 1964	PSSC &	10				17	1.76	.22
	Reg. Phys.	10					1.72	.24
May 1964	PSSC &	12	75	1.43	.46	20	1.78	.14
	Reg. Phys.	10		1.47	.39		1.78	.17
Jan. 1965	PSSC &	10	75	1.35*	.54	20	1.88	.15
	Reg. Phys.	9		1.52	.38		1.88	.10
Jan. 1966	PSSC &	10	75	1.53	.51	22	1.88	.17
	Reg. Phys.	9		1.51	.36		1.88	.14
May 1966	PSSC &	12	75	1.40*	.40	20	1.82	.10
	Reg. Phys.	10		1.58	.33		1.82	.17
Jan. 1967	PSSC &	11	75	1.48	.41	22	1.82	.19
	Reg. Phys.	9		1.59	.34		1.82	.13
May 1967	PSSC &	11	75	1.54	.49	28	1.87	.14
	Reg. Phys.	9		1.59	.36		1.87	.09

*Means significantly different at .05 level

examinees for whom all the relevant data existed. For the Biology Test of May 1963 the reductions in the systematic samples for the BSCS courses were due to lack of verification from teachers that examinees studied the course they said they did as well as unavailability of SAT scores. For several of the studies in chemistry, achievement test performance was adjusted only for performance on subsets of the achievement test rated equally appropriate for both courses and not on SAT; in these instances the systematic samples suffered no reduction due to the unavailability of SAT scores.

In every instance the mean scores in Tables 8, 9, and 10 are for a systematic sample or for a reduced systematic sample. The mean scores are for a systematic sample if no number of examinees is specified for a reduced systematic sample.

For all achievement tests given in May 1963 and thereafter, the significance of the difference between the achievement test mean score of all examinees in a course category (data in Appendix V) and the achievement test mean score of the sample shown in Tables 8, 9, or 10 was determined. The sample means that differed significantly from their population means at the .05 or .01 level are designated by appropriate footnotes in Tables 8, 9, and 10. As discussed in the Design Section an important consideration in each comparison was that the regular course sample mean not differ from the regular course population mean by a significantly different amount than the special course sample mean differed from the special course population mean. Comparisons for which the differences between differences were significant at the .05 and .01 levels are identified by a's and A's, respectively, in Tables 8, 9, and 10.

The data on the actual test performances of the examinees that appears in Tables 8, 9, and 10 will now be presented in graphical form. In Figures 5, 6, and 7 the actual mean scores on the complete science achievement tests of the populations and of the samples of examinees from different courses will be presented. These graphs allow one to identify at a glance the samples that differed markedly from their populations. Figures 8 through 13 pertain only to the samples. In Figures 8, 9, and 10, the actual mean scores of samples of biology examinees are shown. Similar data for chemistry are presented in Figures 11 and 12, and the analogous data for physics appear in Figure 13.

TABLE 8

Actual Performance of Samples of Biology Examinees on the Biology Achievement Tests, on Subsets of Questions in the Biology Achievement Tests Rated Appropriate for Both Courses under Comparison, and on the Scholastic Aptitude Test

Test Form	Test Date	Courses Under Comparison	Number of Examinees in: Course Category	Number of Examinees in: Systematic Sample	Number of Examinees in: Reduced Systematic Sample	Biology Ach. Test: Mean	Biology Ach. Test: SD	Subset of Ach. Test: No. of Questions	Subset of Ach. Test: Mean	Subset of Ach. Test: SD	SAT Verbal: Mean	SAT Verbal: SD	SAT Math: Mean	SAT Math: SD
KAC1	May 1962	BSCS-Blue	58	58	30	506	124	19	9.63	4.2	508	93	516	109
		and Reg. Biol.	10,770	185	154	512	102		9.90	3.7	513	96	510	99
KAC1	May 1962	BSCS-Green	136	136	80	469	97	17	8.80	3.0	453	102	480	118
		and Reg. Biol.	10,770	185	154	512	102		9.84	3.0	513	96	510	99
KAC1	May 1962	BSCS-Yellow	104	104	71	528[1]	112	20	12.24	3.7	504	114	514	125
		and Reg. Biol.	10,770	185	154	512	102		11.39	3.8	513	96	510	99
LAC	May 1963	BSCS-Blue	256	256	112	545	91	22	13.06	3.8	556	96	559	97
		and Reg. Biol.	7,007	185	131	524	102		11.47	4.3	539	89	532	92
LAC	May 1963	BSCS-Green	98	98	36	567[2] A	119	17	10.42	4.2	607	101	596	89
		and Reg. Biol.	7,007	185	131	524	102		9.26	3.6	539	89	532	92
LAC	May 1963	BSCS-Yellow	174	174	53	561	95	20	12.06	4.4	514	98	521	77
		and Reg. Biol.	7,007	185	131	524	102		10.22	4.3	539	89	532	92
OAC1	Jan. 1966	BSCS-Blue	558	465	438	559	101	26	14.10	5.21	549	91	550	96
		and Reg. Biol.	5,480	495	461	526	109		12.72	5.51	533	94	531	93
OAC1	Jan. 1966	BSCS-Green	436	375	357	505	117	17	7.94	4.29	504	116	509	109
		and Reg. Biol.	5,480	495	461	526	109		8.79	3.97	533	94	531	93
OAC1	Jan. 1966	BSCS-Yellow	514	420	411	536	108	30	15.76	6.08	539	98	532	99
		and Reg. Biol.	5,480	495	461	526	109		15.63	6.58	533	94	531	93
OAC2	May 1966	BSCS-Blue	1,262	495	463	575	100	22	13.25	5.00	554	102	555	102
		and Reg. Biol.	6,203	-	499	548[1]	103		11.71	5.25	526	98	530	103
OAC2	May 1966	BSCS-Green	1,022	495	403	520	104	21	9.75	4.95	500	101	512	105
		and Reg. Biol.	6,203	-	499	548[1a]	103		11.14	4.96	526	98	530	103

TABLE 8 (con't)

Test Form	Test Date	Courses Under Comparison	Number of Examinees in: Course Category	Systematic Sample	Reduced Systematic Sample	Biology Ach. Test: Mean	Biology Ach. Test: SD	Subset of Ach. Test: No. of Questions	Subset of Ach. Test: Mean	Subset of Ach. Test: SD	SAT Verbal: Mean	SAT Verbal: SD	SAT Math: Mean	SAT Math: SD
OAC2	May 1966	BSCS-Yellow	2,375	-	497	551	98	23	14.24	4.95	527	99	527	102
		and Reg. Biol.	6,203	-	499	548[1a]	103		13.81	5.09	526	98	530	103
PAC1	Jan. 1967	BSCS-Blue	798	-	500	537	101	34	16.45	7.00	547	97	549	103
		and Reg. Biol.	4,855	-	498	512	110		13.58	7.39	534	101	534	102
PAC1	Jan. 1967	BSCS-Green	669	-	500	498	111	32	13.84	7.19	509	109	516	102
		and Reg. Biol.	4,855	-	498	512	110		14.32	6.94	534	101	534	102
PAC1	Jan. 1967	BSCS-Yellow	1,091	-	496	526	109	36	16.14	7.82	533	99	538	100
		and Reg. Biol.	4,855	-	498	512	110		14.40	7.86	534	101	534	102
PAC2	May 1967	BSCS-Blue	1,588	-	500	556	103	32	19.19	6.58	545	103	557	101
		and Reg. Biol.	5,207	-	499	526	110		16.24	7.01	519	105	524	100
PAC2	May 1967	BSCS-Green	1,213	-	500	518	108	39	18.82	7.93	508	106	518	109
		and Reg. Biol.	5,207	-	499	526	110		19.12	8.04	519	105	524	100
PAC2	May 1967	BSCS-Yellow	2,848	-	500	548[3]	106	50	27.31	10.06	526	104	540	106
		and Reg. Biol.	5,207	-	499	526	110		24.37	10.36	519	105	524	100

[1]Sample mean significantly higher than population mean at .05 level

[2]Sample mean significantly higher than population mean at .01 level

[3]Sample mean significantly lower than population mean at .01 level

[a]Difference between regular sample and population means significantly different from difference between special sample and population means at .05 level

[A]Difference between regular sample and population means significantly different from difference between special sample and population means at .01 level

TABLE 9

Actual Performance of Samples of Chemistry Examinees on the Chemistry Achievement Tests, on Subsets of Questions in the Chemistry Achievement Tests Rated Appropriate for Both Courses under Comparison, and on the Scholastic Aptitude Test

			Number of Examinees in			Chemistry Ach. Test		Subset of Ach. Test			SAT Verbal		SAT Math	
Test Form	Test Date	Courses Under Comparison	Course Category	Systematic Sample	Reduced Systematic Sample	Mean	SD	No. of Questions	Mean	SD	Mean	SD	Mean	SD
IAC	Dec. 1961	CBA &	-	64	-	537	111	76	35.88		-	-	-	-
		Reg. Chem.	-	276	-	561	109		38.00		-	-	-	-
IAC	Dec. 1961	CHEMS &	-	32	-	553	130	74	39.34	16.43	-	-	-	-
		Reg. Chem.	-	276	-	561	109		39.04	13.91	-	-	-	-
K-FAC	Mar. 1962	CBA &	-	107	-	498	92	67	30.77	10.69	-	-	-	-
		Reg. Chem.	-	426	-	532	106		33.85	12.06	-	-	-	-
K-FAC	Mar. 1962	CHEMS &	-	52	-	492	94	67	29.75	11.65	-	-	-	-
		Reg. Chem.	-	426	-	532	106		32.90	12.36	-	-	-	-
KAC1	May 1962	CBA &	-	310	-	546	95	71	34.17	12.49	-	-	-	-
		Reg. Chem.	-	350	-	559	95		32.81	12.28	-	-	-	-
KAC1	May 1962	CHEMS &	-	405	-	527	85	70	32.14	11.79	-	-	-	-
		Reg. Chem.	-	350	-	559	95		32.71	12.34	-	-	-	-
LAC	May 1963	CBA &	560	370	-	555	102	20	12.15	5.04	-	-	-	-
		Reg. Chem.	11,611	370	-	581	108		11.78	5.26	-	-	-	-
LAC	May 1963	CHEMS &	1,472	370	-	547	106	20	12.15	5.27	-	-	-	-
		Reg. Chem.	11,611	370	-	581	108		11.78	5.26	-	-	-	-
MAC2	May 1964	CBA &	878	365	-	525	95	20	10.02	4.42	-	-	-	-
		Reg. Chem.	12,414	225	-	577	104		11.04	4.26	-	-	-	-
MAC2	May 1964	CHEMS &	3,495	369	-	534	97	30	14.49	6.39	-	-	-	-
		Reg. Chem.	12,414	225	-	577	104		15.24	6.50	-	-	-	-

TABLE 9 (con't)

Test Form	Test Date	Courses Under Comparison	Number of Examinees in: Course Category	Systematic Sample	Reduced Systematic Sample	Chemistry Ach. Test: Mean	SD	Subset of Ach. Test: No. of Questions	Mean	SD	SAT: Verbal Mean	Verbal SD	Math Mean	Math SD
NAC1	Jan. 1965	CBA & Reg. Chem.	517 6,342	370 370	358 358	546 572[1A]	108 111	20	10.46 10.74	4.89 4.79	543 559	110 102	615 629	107 97
NAC1	Jan. 1965	CHEMS & Reg. Chem.	1,731 6,342	370 370	337 358	574 572[1A]	103 111	20	11.76 10.74	4.43 4.79	558 559	92 102	633 629	94 97
NAC2	May 1965	CBA & Reg. Chem.	1,093 11,585	364 370	- -	547 581	93 98	20	10.22 11.40	3.98 4.06	- -	- -	- -	- -
NAC2	May 1965	CHEMS & Reg. Chem.	5,007 11,585	370 370	- -	573 581	96 98	20	11.17 11.40	4.18 4.06	- -	- -	- -	- -
OAC1	Jan. 1966	CBA & Reg. Chem.	527 5,057	500 500	- -	535 575	117 106	26	13.07 14.83	5.95 5.70	- -	- -	- -	- -
OAC1	Jan. 1966	CHEMS & Reg. Chem.	2,295 5,057	500 500	- -	568 575	116 106	29	16.40 16.17	6.47 6.06	- -	- -	- -	- -
OAC2	May 1966	CBA & Reg. Chem.	1,121 9,894	- -	496 497	545 563	101 101	19	10.41 10.91	4.20 4.13	532 549	104 92	598 601	101 89
OAC2	May 1966	CHEMS & Reg. Chem.	6,537 9,894	- -	493 497	562 563	101 101	31	16.36 16.50	6.36 6.34	542 549	96 92	600 601	99 89
PAC1	Jan. 1967	CBA & Reg. Chem.	510 4,371	500 -	439 500	568 583	103 102	38	17.59 19.36	9.14 9.20	534 558	109 98	610 618	102 93
PAC1	Jan. 1967	CHEMS & Reg. Chem.	2,833 4,371	- -	500 500	595 583	104 102	44	22.40 22.63	10.16 10.35	548 558	101 98	628 618	95 93
PAC2	May 1967	CBA & Reg. Chem.	964 8,215	- -	500 500	532[2] 565	111 109	31	13.22 14.68	6.58 6.39	528 551	107 102	584 603	107 97
PAC2	May 1967	CHEMS & Reg. Chem.	8,121 8,215	- -	500 500	576 565	106 109	45	23.89 23.45	8.05 8.70	546 551	96 102	608 603	95 97

[1]Sample mean significantly higher than population mean at .01 level

[2]Sample mean significantly lower than population mean at .05 level

[A]Difference between Reg. sample and population means significantly different from difference between special sample and population means at .01 level

TABLE 10

Actual Performance of Samples of Physics Examinees on the Physics Achievement Tests, on Subsets of Questions in the Physics Achievement Tests Rated Appropriate for Both Courses under Comparison, and on the Scholastic Aptitude Test

			Number of Examinees in			Physics Ach. Test		Subset of Ach. Test			SAT Verbal		SAT Math	
Test Form	Test Date	Courses Under Comparison	Course Category	Systematic Sample	Reduced Systematic Sample	Mean	SD	No. of Questions	Mean	SD	Mean	SD	Mean	SD
LAC2	May 1963	PSSC & Reg. Phys.	2,121 4,456	370 370	369 359	592 563	98 92	19	8.54 7.36	3.90 3.77	586 565	98 98	662 632	80 96
MAC1	Jan. 1964	PSSC & Reg. Phys.	2,036 3,825	370 370	355 363	558 556	99 92	17	6.55 5.93	3.79 3.79	565 549	93 100	660 640	81 90
MAC2	May 1964	PSSC & Reg. Phys.	2,608 3,694	370 370	330 337	579 561	84 83	20	10.26 9.56	4.18 3.99	600 580	90 97	680 656	75 85
NAC1	Jan. 1965	PSSC & Reg. Phys.	2,640 4,006	370 370	313 332	575 576	96 96	20	8.63 8.58	4.50 4.47	562 554	96 97	652 634	84 89
OAC1	Jan. 1966	PSSC & Reg. Phys.	2,636 3,499	495 495	476 457	590 582	96 96	22	10.50 10.38	4.79 4.69	560 553	100 100	645 629	86 89
OAC2	May 1966	PSSC & Reg. Phys.	2,504 2,685	- -	514 538	630 621[1]	96 95	20	11.58 10.91	3.82 4.03	582 563	93 93	661 637	84 90
PAC1	Jan. 1967	PSSC & Reg. Phys.	3,067 3,239	- -	500 498	584 569	94 101	22	8.79 7.63	4.70 4.88	548 538	99 101	638 619	89 92
PAC2	May 1967	PSSC & Reg. Phys.	2,629 2,651	- -	500 500	607[2] 582[3]	96 100	28	15.82 13.23	5.73 6.04	566 549	103 108	649 623	91 96

[1]Sample mean significantly higher than population mean at .01 level

[2]Sample mean significantly lower than population mean at .01 level

[3]Sample mean significantly lower than population mean at .05 level

Samples were drawn and analyses of test appropriateness were done for Biology Tests given on six of the test dates listed in Figure 5; these six dates were May 1962, May 1963, January 1966, May 1966, January 1967, and May 1967. The performance of four samples of students on each of these six tests was analyzed. Twenty of the 24 samples means did not differ significantly from their population means. The four sample means that did differ significantly from their population means were as follows:

1. May 1962, BSCS-Yellow, Sample mean significantly higher than population mean at .05 level
2. May 1963, BSCS-Green, Sample mean significantly higher than population mean at .01 level
3. May 1966, Regular, Sample mean significantly higher than population mean at .05 level
4. May 1967, BSCS-Yellow, Sample mean significantly lower than population mean at .01 level

For two comparisons (BSCS-Green and regular, May 1966 and BSCS-Yellow and regular, May 1966) the differences between regular sample and population means were significantly different from the differences between BSCS sample and population means at the .05 level. The one comparison that will be ruled out of consideration because the difference between the regular sample and population means is significantly different from the difference between the BSCS sample and population mean at the .01 level sticks out plainly in Figure 5. It is the comparison between Regular Biology and BSCS-Green in May 1963.

Samples were drawn and analyses of test appropriateness were done for Chemistry Tests given on eight of the test dates shown in Figure 6. January 1964 was the single omission. Twenty-two of the sample means did not differ significantly from their population means. The two sample means that did differ significantly from their population means were as follows:

1. January 1965, Regular, Sample mean significantly higher than population mean at .01 level
2. May 1967, CBA, Sample mean significantly lower than population mean at .05 level

Figure 5

Biology - Mean Scores on the Complete Biology Achievement Tests of Populations and Samples of Examinees from Different Biology Courses

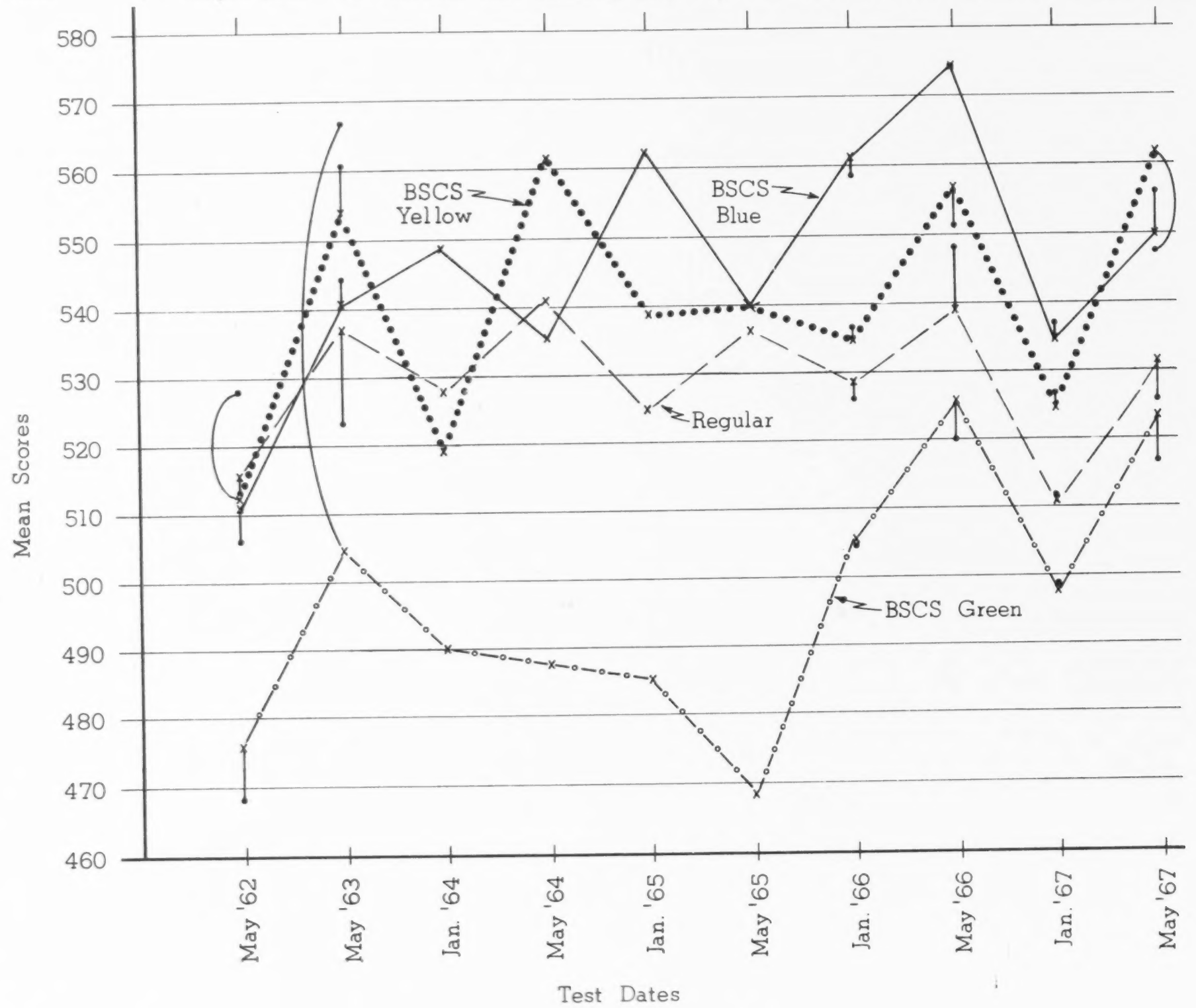
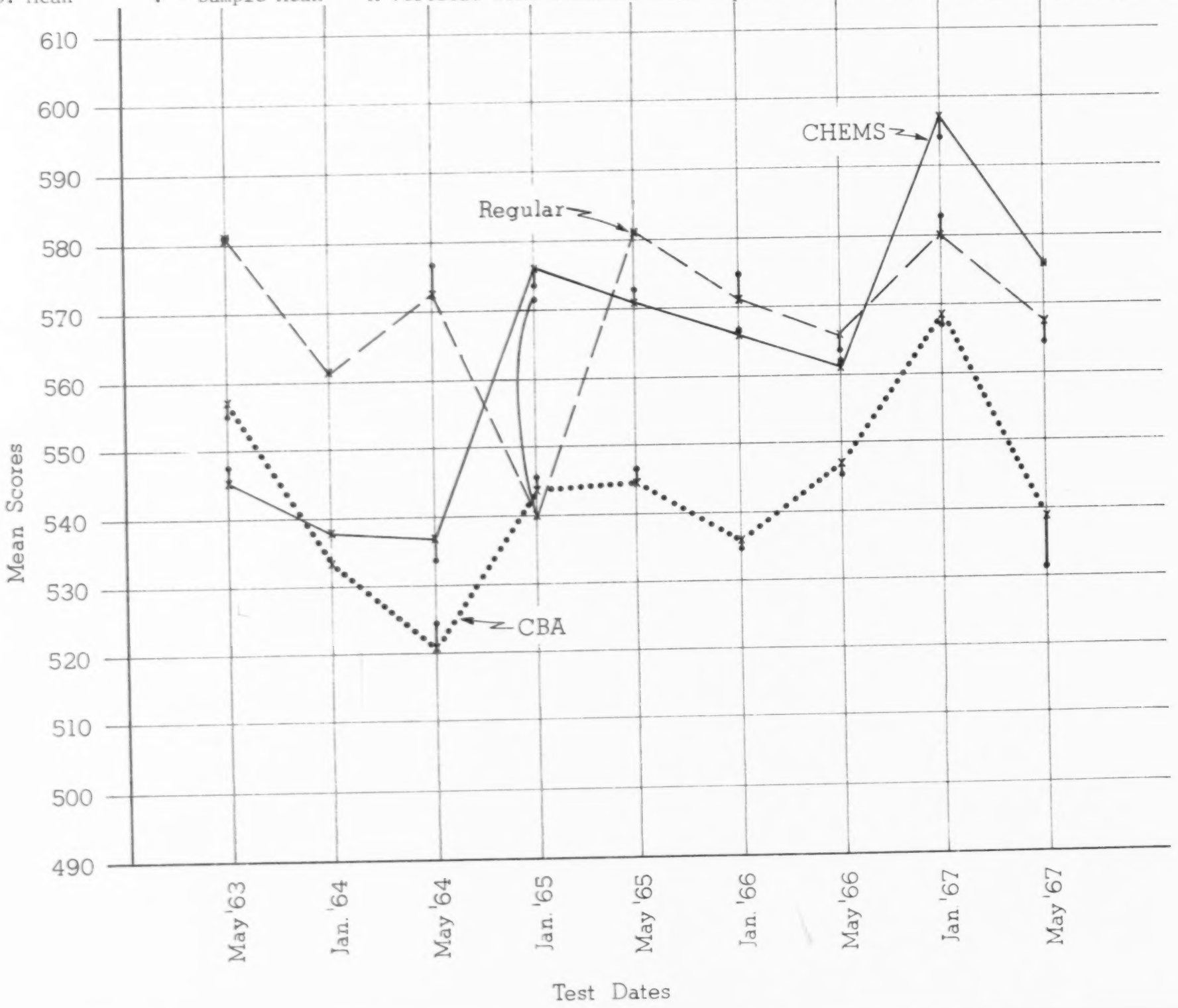


Figure 6

Chemistry - Mean Scores on the Complete Chemistry Achievement Tests of Populations and Samples of Examinees from Different Chemistry Courses

x - Pop. Mean . - Sample Mean A vertical line connects each Pop. Mean with its corresponding Sample Mean



The two comparisons that will be ruled out of consideration because differences between Regular Chemistry sample and population means are significantly different from differences between special course sample and population means are immediately apparent in Figure 6. They are the comparisons between CBA and regular and between CHEMS and regular in January 1965.

Samples were drawn and analyses of test appropriateness were done for Physics Tests given on all of the test dates shown in Figure 7 except May 1965. Three of the 16 sample means did differ significantly from their population means. They were as follows:

1. May 1966, Regular, Sample mean significantly higher than population mean at .01 level
2. May 1967, PSSC, Sample mean significantly lower than population mean at .01 level
3. May 1967, Regular, Sample mean significantly lower than population mean at .05 level

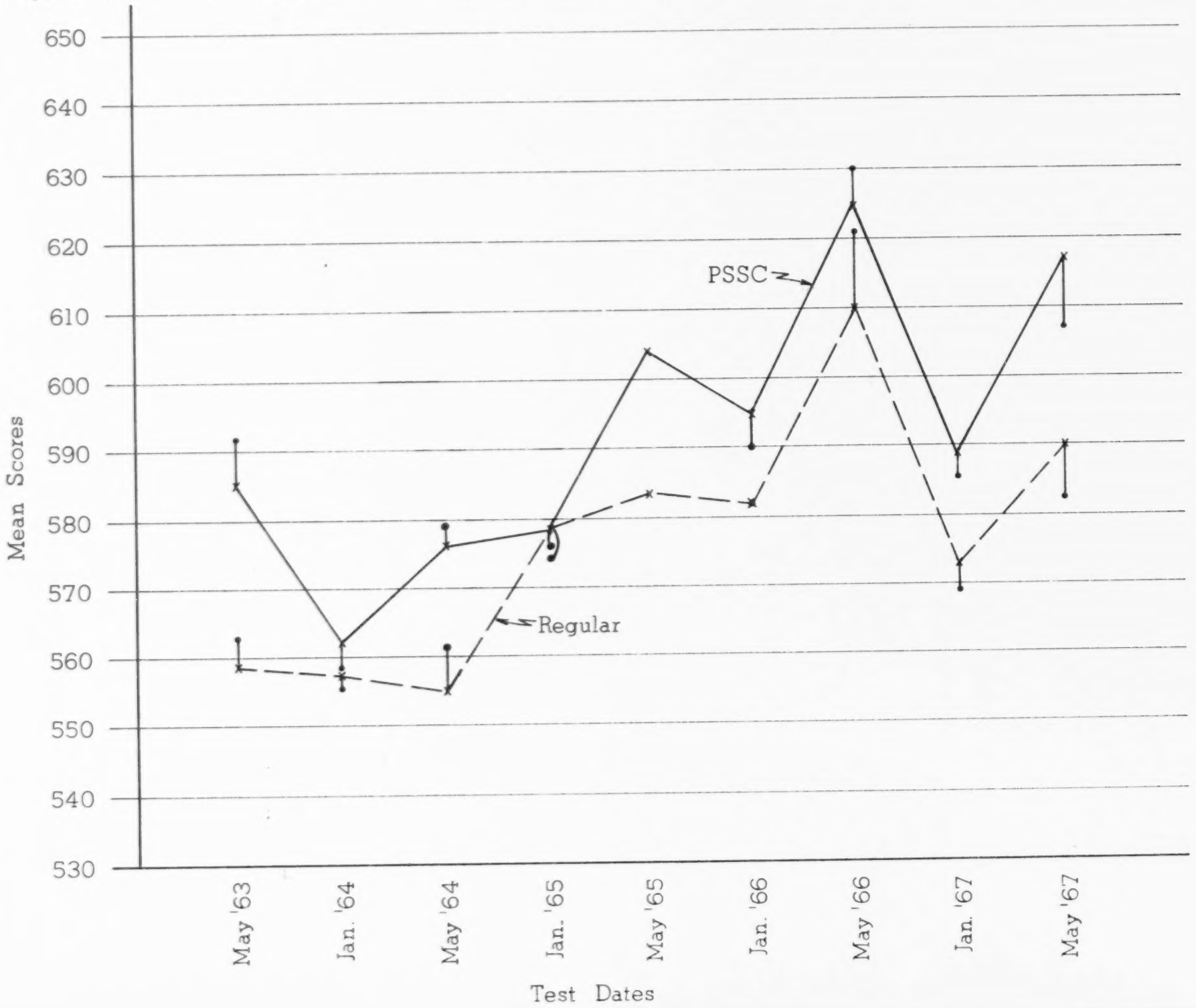
Fortunately, in each comparison the difference between the regular sample and population mean was nearly the same as the difference between the PSSC sample population mean. None of the differences differ significantly at the .01 level; none of the comparisons need be ruled out.

In Figures 8 through 13 the actual mean scores of samples of examinees from different courses are shown. In any one figure the mean scores of examinees from only two courses are arrayed. One of those courses is the regular course, and dashed

Figure 7

Physics - Mean Scores on the Complete Physics Achievement Tests of Populations and Samples of Examinees from Different Physics Courses

x- Pop. Mean . - Sample Mean A vertical line connects each Pop. Mean with its corresponding Sample Mean



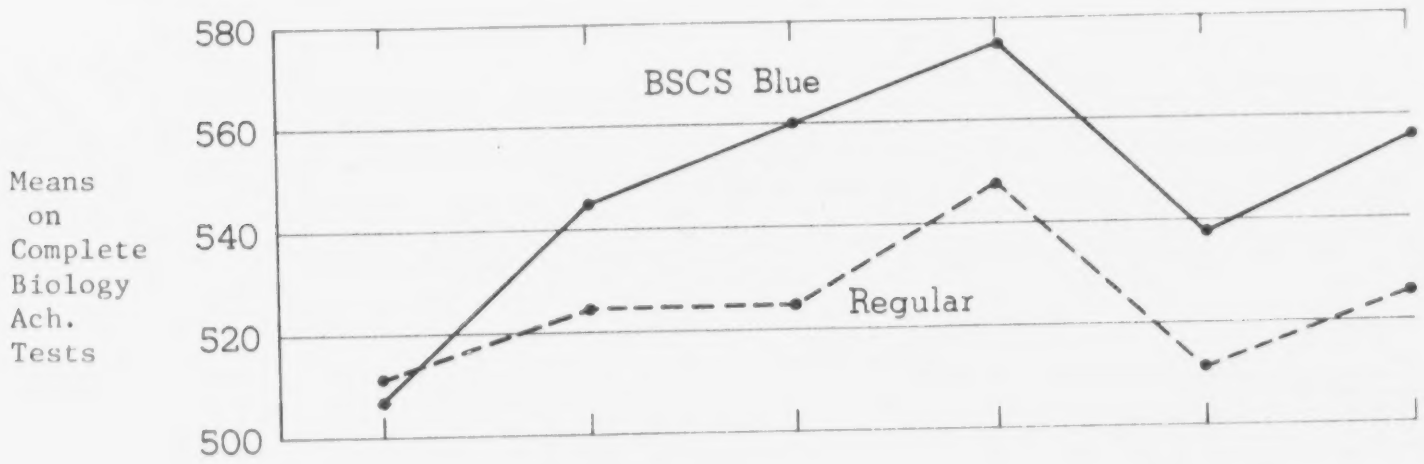
lines are used throughout to indicate the mean scores of the regular-course examinees. Attention is directed to the differences between mean scores of samples on complete achievement tests and the corresponding differences between the means of the same samples on the subtest and on SAT-V and M. For example, if course X examinees score higher than course Y examinees on a complete achievement test, one would expect course X examinees to score higher on the appropriate-for-both subtest and on SAT-V and M, also. If they do, there is evidence that some or all of the superior performance of the X examinees on the achievement test can be ascribed to their superior science achievement and scholastic aptitude. On the other hand, if X examinees score higher on the achievement test, but lower on the appropriate-for-both subtests and on SAT, there is evidence of achievement test bias in favor of course X. The following figures are a useful first approach to the differing performances of examinees from difference courses on achievement tests and related measures. No particular significance should be attached to the absolute value of the mean scores on the appropriate-for-both subtests. These subtests had widely differing numbers of items and all the subtest scores are certainly not on the same scale.

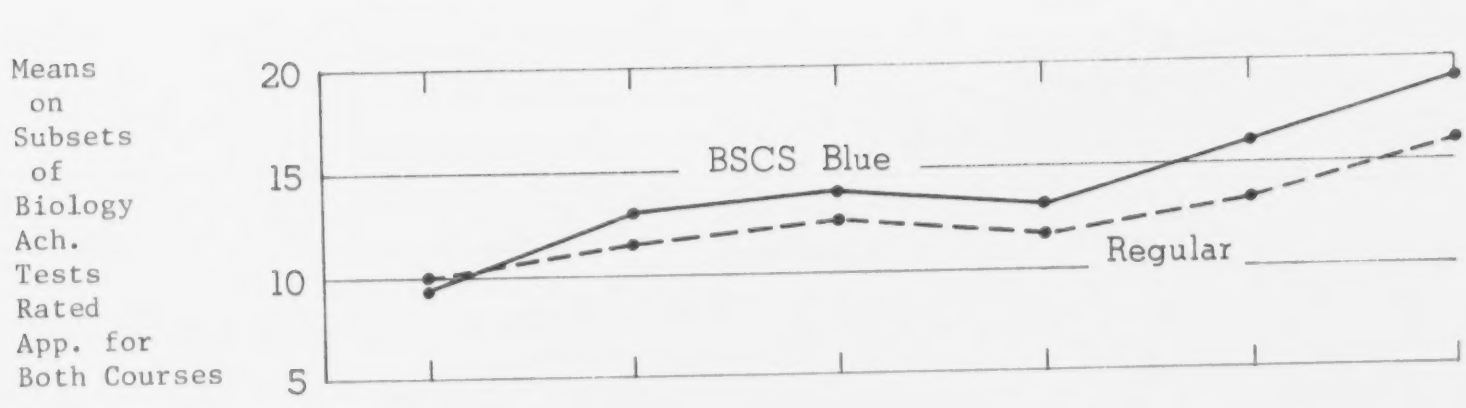
The data in Figure 8 show that, except for the May 1962 Biology Test, the BSCS-Blue examinees score consistently higher than the regular examinees on the complete Biology Achievement Tests, on the appropriate-for-both subtests, and on SAT-V and M.

For all six of the tests shown in Figure 9 the examinees that score higher on the achievement tests also score higher on the appropriate-for-both subtests and on SAT-V and M. For five test dates the regular examinees score higher than the BSCS-Green examinees on all measures. The BSCS-Green sample of May 1963 was very atypical and not representative of its population. The sample mean was 567 whereas the population mean was 505. The sample mean was significantly higher than the population mean at the .01 level. Furthermore, the difference between the BSCS-Green sample and population means was significantly different from the difference between the Regular Biology sample and population mean at the

Figure 8

Biology: BSCS-Blue and Regular Sample Means on the Complete Biology Achievement Tests, on Subsets of the Biol. Ach. Tests Rated Appropriate for Both Courses, and on SAT - V and M





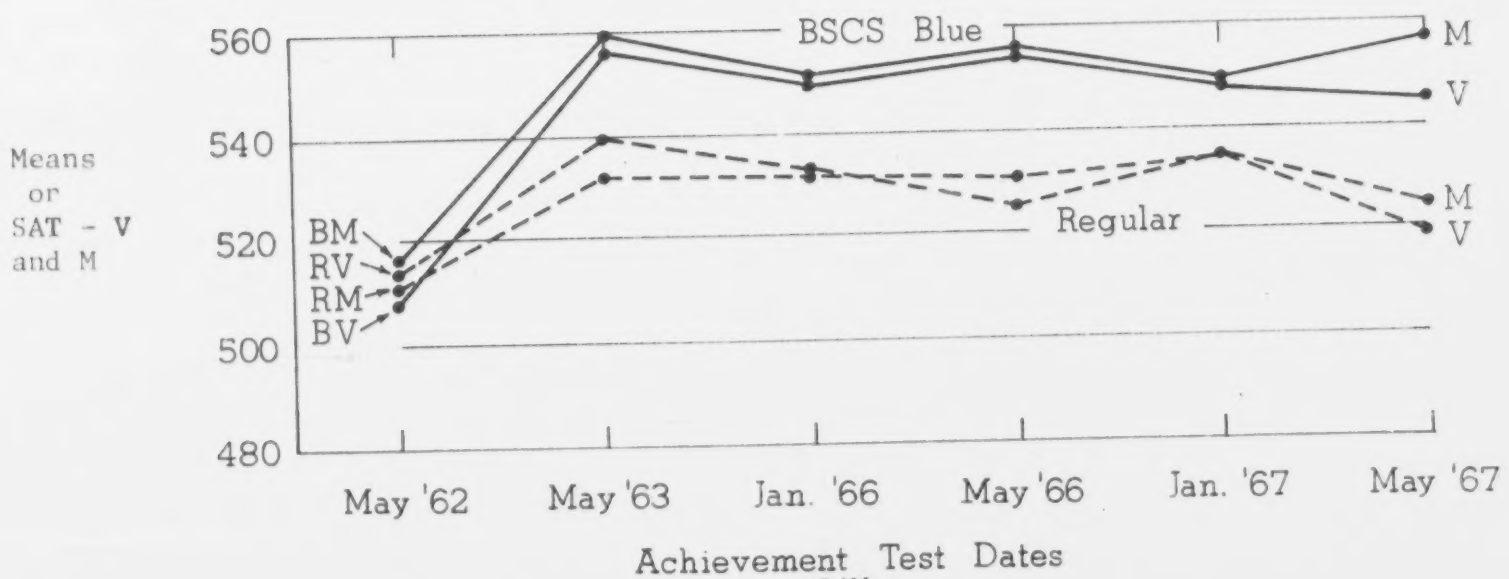
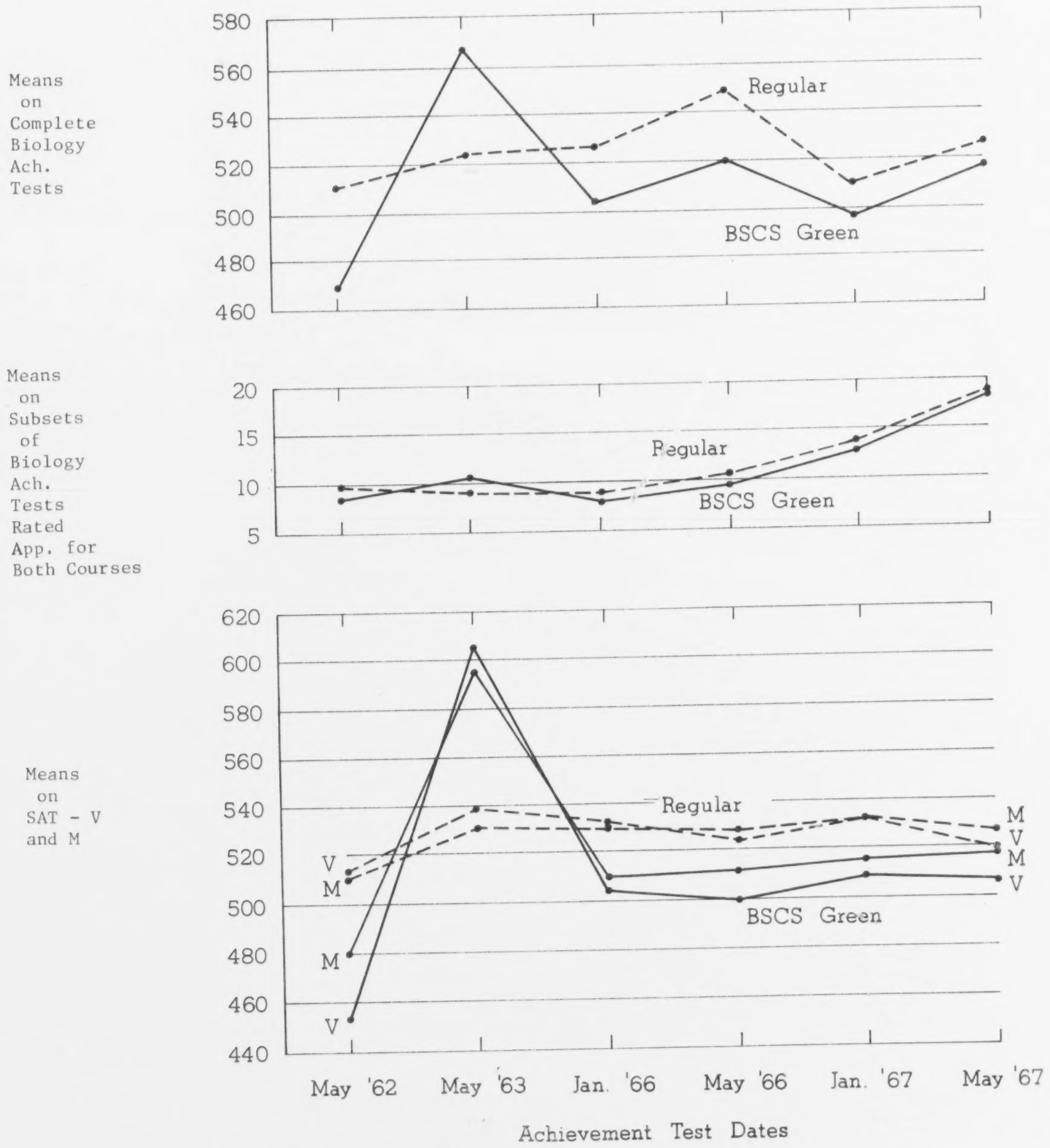


Figure 9

Biology: BSCS-Green and Regular Sample Means on the Complete Biology Achievement Tests, on Subsets of the Biol. Ach. Tests Rated Appropriate for Both Courses, and on SAT - V and M



.01 level. The comparison between BSCS-Green and regular in May 1963 will not be analyzed or considered further.

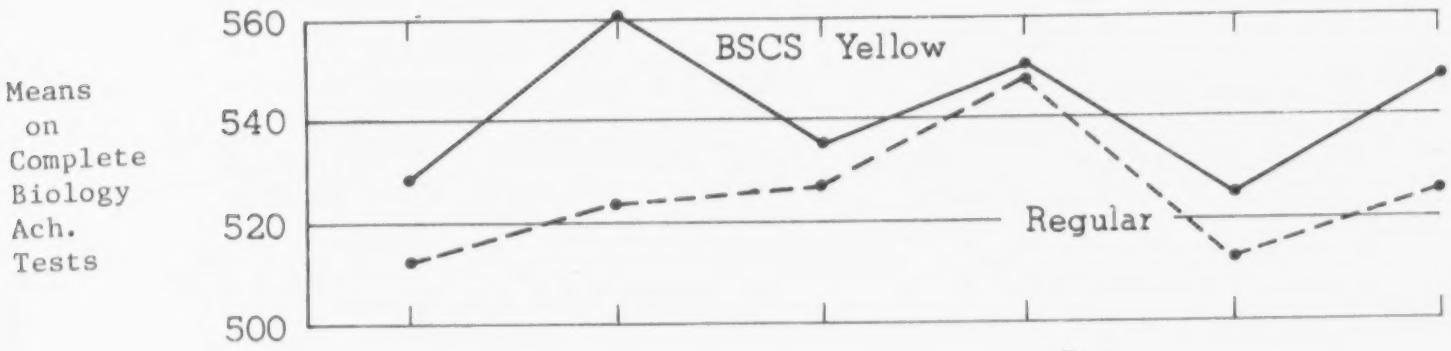
On all six test dates in Figure 10, the BSCS-Yellow examinees score higher than the regular examinees on the achievement tests and on the appropriate-for-both subtests. The SAT means are closely clustered, but in May 1963 the regular examinees scored higher than the BSCS-Yellow examinees on SAT, suggesting possible bias in this achievement test in favor of BSCS-Yellow students.

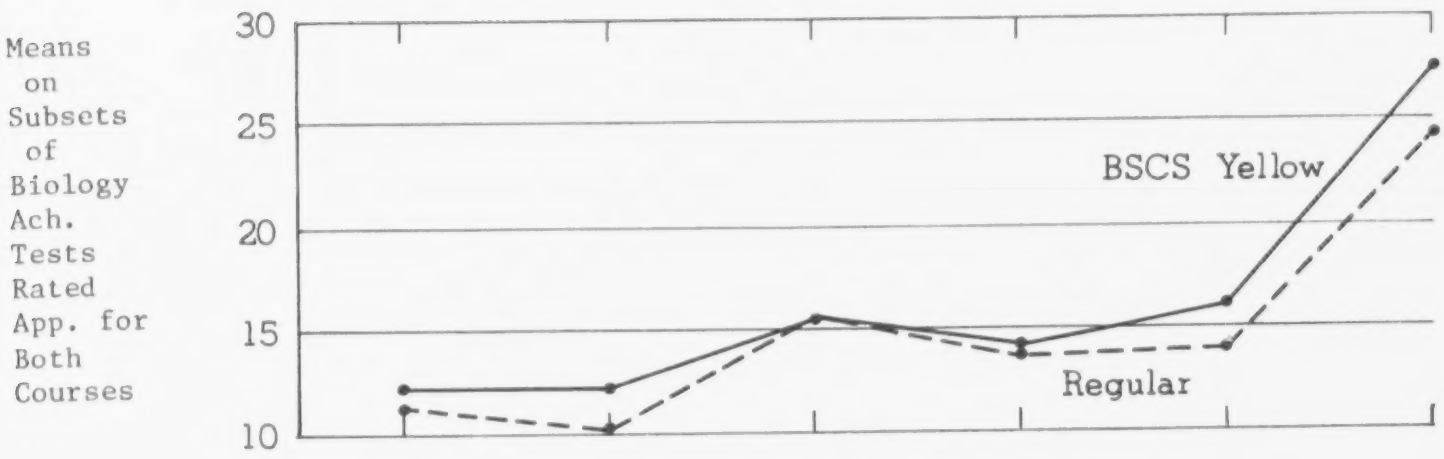
In Figures 11 and 12 only chemistry data from May 1963 and later are presented although earlier data are given in Table 9. The reason is that the chemistry investigations prior to May 1963 did not involve ratings from teachers of regular courses; the test questions were simply assumed to be appropriate for regular students. Beginning with the May 1963 test the appropriate-for-both subtests were identified on the basis of ratings for both regular and special courses. The result of this change was an appreciable drop in the number of questions in the appropriate-for-both subtests.

For January 1965 in chemistry the difference between the CBA sample and population means and the difference between the CHEMS sample and population means were both significantly different from the difference between the regular sample and population means at the .01 level. Neither chemistry comparison for January 1965 will be analyzed or considered further.

Figure 10

Biology: BSCS-Yellow and Regular Sample Means on the Complete Biology Achievement Tests, on Subsets of the Biol. Ach. Tests Rated Appropriate for Both Courses, and on SAT - V and M





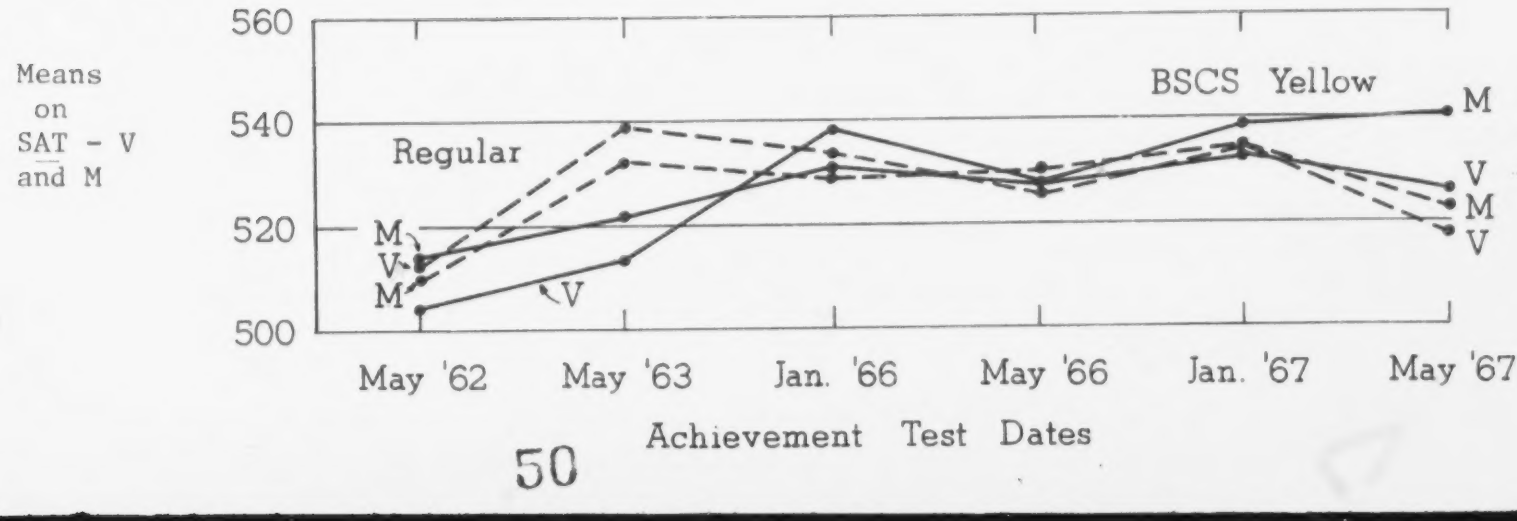
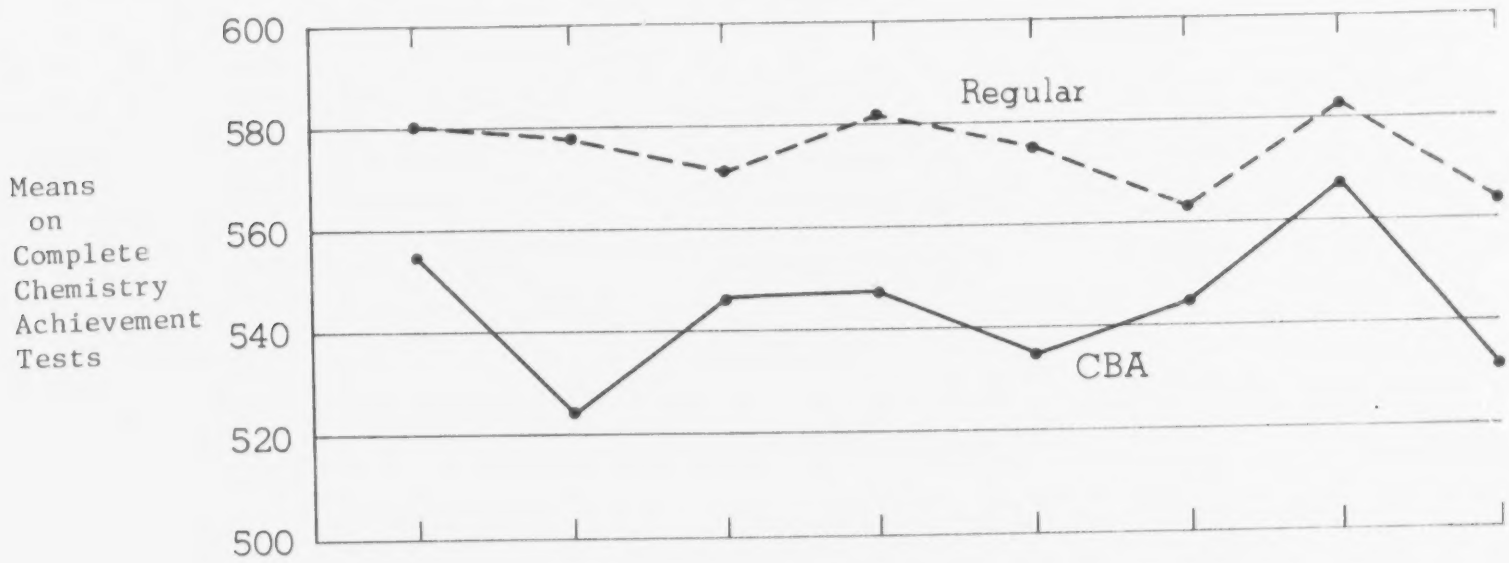
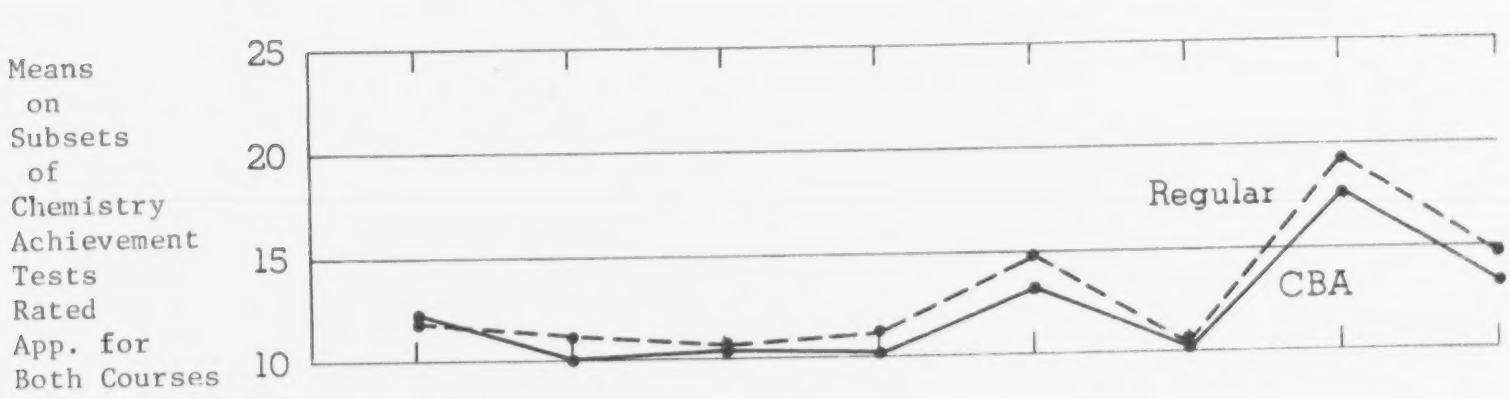


Figure 11

Chemistry: CBA and Regular Sample Means on the Complete Chemistry Achievement Tests, on Subsets of the Chem. Ach. Tests Rated Appropriate for Both Courses, and on SAT - V and M





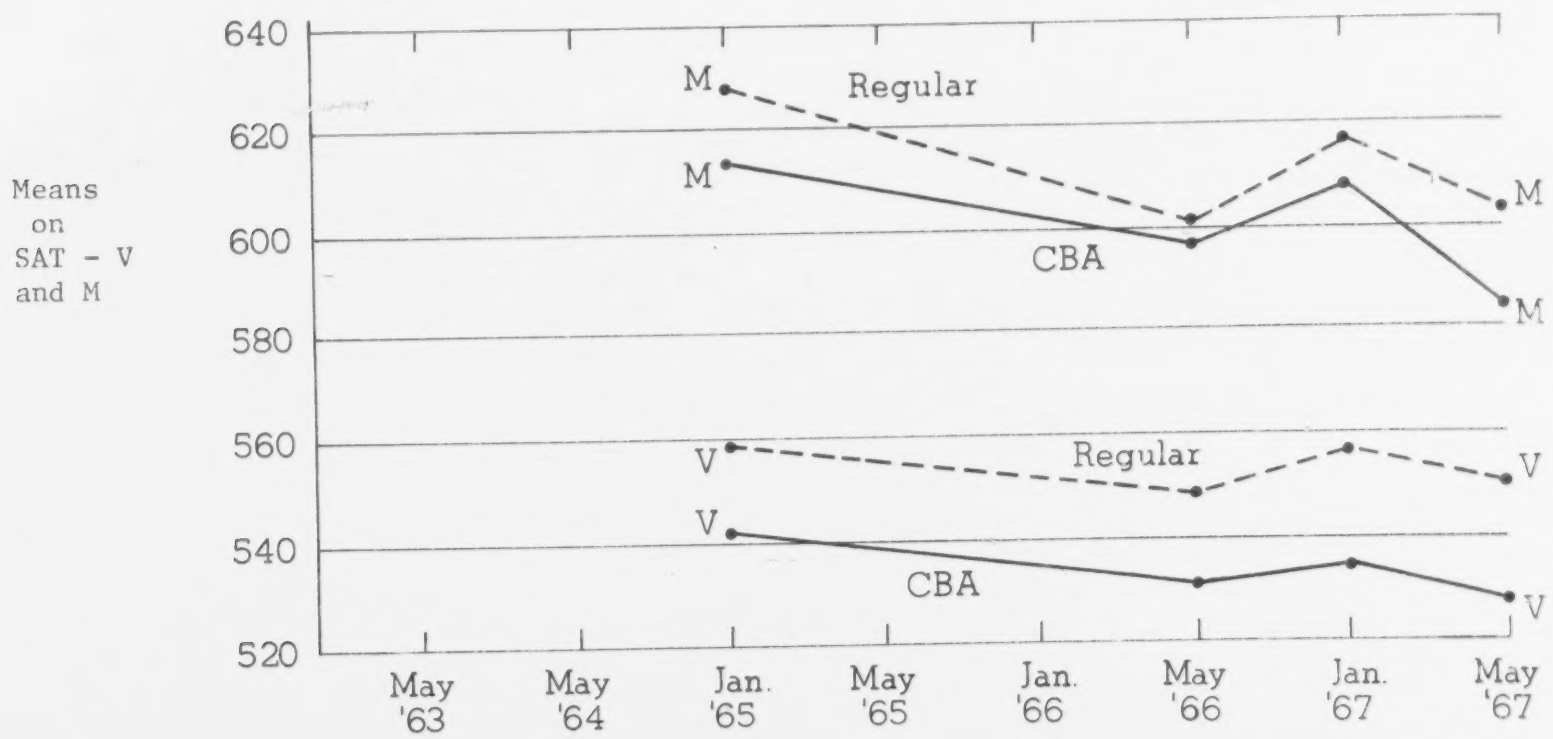


Figure 12

Chemistry: CHEMS and Regular Sample Means on the Complete Chemistry Achievement Tests, on Subsets of the Chem. Ach. Tests Rated Appropriate for Both Courses, and on SAT - V and M

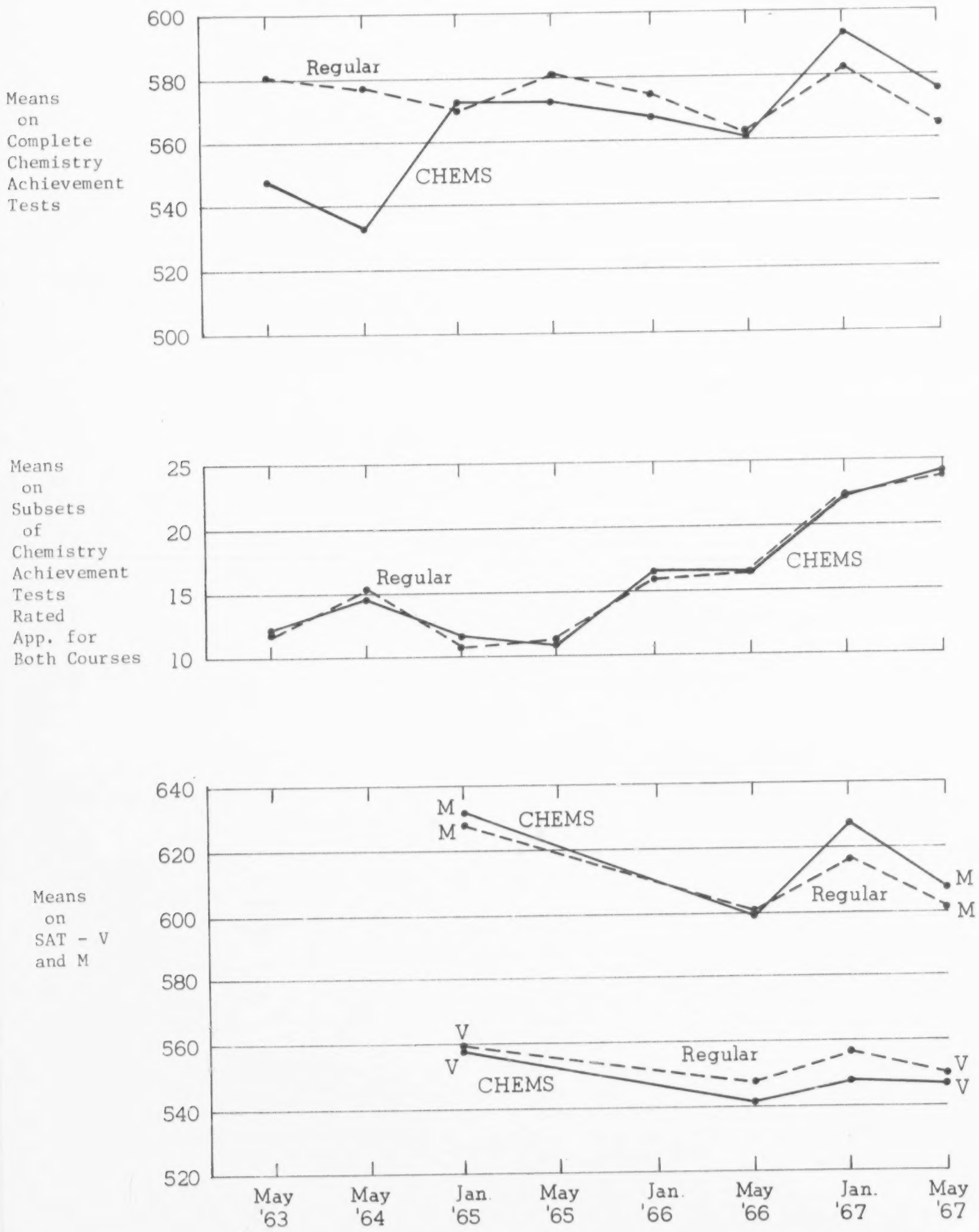
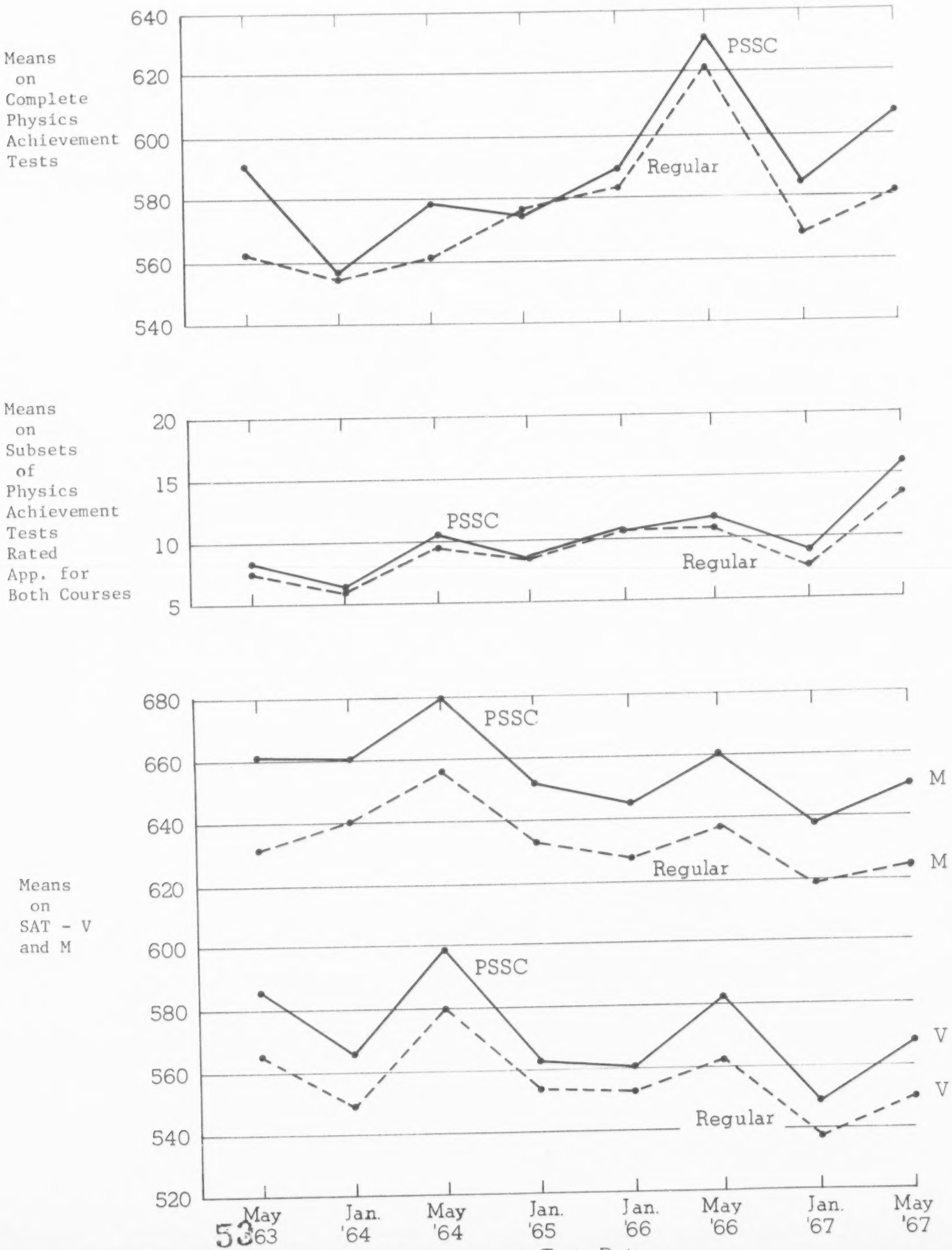


Figure 13

Physics: PSSC and Regular Sample Means on the Complete Physics Achievement Tests, on Subsets of the Physics Ach. Tests Rated Appropriate for Both Courses, and on SAT - V and M



A major feature of the actual performance data presented in Tables 8-10 and depicted for the most part in Figures 8-13 is that the students from the course in each comparison who score higher on the achievement test almost always score higher on the appropriate-for-both items and on SAT-V and M. If one takes as the indication of bias that one course in a comparison have the higher mean on the achievement test and the other course have the higher mean(s) on all available concomitant measures, the data can be summarized as shown in the chart below.

Subject	Comparison	Comment
Biology	BSCS-Blue & Regular	No indications of bias
	BSCS-Green & Regular	No indications of bias
	BSCS-Yellow & Regular	No indications of bias
Chemistry	CBA and Regular	There are indications of bias in favor of regular students on two tests, May 1962 and May 1963.
	CHEMS and Regular	On three tests, December 1961, May 1963, and January 1966, there are indications of bias in favor of regular students.
Physics	PSSC and Regular	There is an indication of bias in favor of regular students on the January 1965 test.

For the January and May 1967 Biology, Chemistry, and Physics Tests, checks on the accuracy of student responses regarding courses studied were made. These checks involved asking schools to verify or deny the responses of their students regarding the courses they studied. In Table 11 the test performances in January and May 1967 of student-response samples and of school-verified samples are presented. Each school-verified sample is a subsample of the corresponding student-response sample.

The data in Table 11 are from 18 student-response samples and 18 corresponding school-verified samples; there are eight samples of each kind in biology, six in chemistry, and four in physics. In 17 of the 18 cases, the school-verified sample had a higher mean score on the achievement test than the corresponding student-response sample. The lone exception is the Regular Biology sample of January 1967.

TABLE 11

Actual Performance of Student-Response and School-Verified Samples of Biology, Chemistry, and Physics Examinees from January and May 1967 on the Biology, Chemistry, and Physics Achievement Tests, respectively; on Subsets of the Biology, Chemistry, and Physics Tests Rated Appropriate for Courses under Comparison, respectively; and on the Scholastic Aptitude Test

Subject	Ach. Test Date	Courses Under Comparison	Number of Examinees in:		Mean Scores							
					Complete Ach. Test		Subset of Ach. Test		SAT-V		SAT-M	
			Stud. Resp. Sample	Sch. Ver. Sample	Stud. Resp. Sample	Sch. Ver. Sample	Stud. Resp. Sample	Sch. Ver. Sample	Stud. Resp. Sample	Sch. Ver. Sample	Stud. Resp. Sample	Sch. Ver. Sample
Biology	Jan. '67	BSCS-Blue & Reg. Biol.	500 498	378 271	537 512	544[2] 509	16.45 13.58	16.88 13.32	547 534	557 531	549 534	558 529
	Jan. '67	BSCS-Green & Reg. Biol.	500 498	342 271	498 512	517[1] 509[a]	13.84 14.32	14.85 14.16	509 534	534 531	516 534	533 529
	Jan. '67	BSCS-Yellow & Reg. Biol.	496 498	370 271	526 512	528 509	16.14 14.40	16.24 14.10	533 534	534 531	538 534	541 529
	May '67	BSCS-Blue & Reg. Biol.	500 499	420 314	556 526	560[2] 529	19.19 16.24	19.57 16.42	545 519	549 521	557 524	561 527
	May '67	BSCS-Green & Reg. Biol.	500 499	375 314	518 526	538[1] 529[a]	18.82 19.12	20.17 19.43	508 519	529 521	518 524	537 527
	May '67	BSCS-Yellow & Reg. Biol.	500 499	432 314	548[4] 526	549[4] 529	27.31 24.37	27.37 24.67	526 519	531 521	540 524	541 527
Chemistry	Jan. '67	CBA & Reg. Chem.	439 500	260 404	568 583	585[1] 586	17.59 19.36	18.64 19.72	534 557	552 562	610 618	630 620
	Jan. '67	CHEMS & Reg. Chem.	500 500	371 404	595 583	608[2] 586	22.40 22.63	23.32 23.02	548 557	560 562	628 618	640 620
	May '67	CBA & Reg. Chem.	500 500	307 415	532[3] 565	554[1] 569	13.22 14.68	14.38 15.01	528 551	550 554	584 603	606 608
	May '67	CHEMS & Reg. Chem.	500 500	403 415	576 565	582 569	23.89 23.45	24.34 23.79	546 551	554 554	608 603	615 608
Physics	Jan. '67	PSSC & Reg. Phys.	500 498	404 392	584 569	591 572	8.79 7.63	9.20 7.80	548 538	556 543	638 619	645 622
	May '67	PSSC & Reg. Phys.	500 500	392 392	607[4] 582[3]	619 584	15.82 13.23	16.60 13.32	566 549	573 553	649 623	658 629

[1]Sample mean significantly higher than population mean at .01 level

[2]Sample mean significantly higher than population mean at .05 level

[3]Sample mean significantly lower than population mean at .05 level

[4]Sample mean significantly lower than population mean at .01 level

[a]Difference between BSCS-Green sample and population means significantly different from difference between Reg. Biol. sample and population means at .05 level

In all 18 cases the differences in mean scores on the complete achievement test of student-response and school-verified samples are accompanied by differences in the same direction in the mean scores on the appropriate-for-both subtest and on SAT-V and M.

Considering the 12 comparisons based on school-verified samples, for six of six comparisons in biology and both comparisons in physics higher means on the science achievement tests are associated with higher means on all three concomitant variables. This same pattern holds in chemistry for the appropriate-for-both questions, but there are some exceptions in the relationship between performance on the Chemistry Achievement Test and SAT.

E. Adjusted Performances of Examinees in Different Courses on the Science Achievement Tests After Taking Account of Their Performances on Subsets of the Achievement Tests Rated Appropriate for Different Courses and on SAT

The mean scores on the science achievement tests expected of students in the regular and special courses who were equivalent in performance on the appropriate-for-both subsets and on SAT will be presented in Tables 12, 13, and 14 for biology, chemistry, and physics, respectively. The most valid adjustments and comparisons are for those instances where there were no significant differences between the two courses under comparison in terms of three considerations: (1) The difference between the regular-course population and sample means not significantly different from the difference between the special-course population and sample means; (2) the regression-line slope for the regular course not significantly different from the regression-line slope for the special course; (3) the standard error of estimate for the regular course not significantly different from the standard error of estimate for the special course. Significant differences in each of these three categories at the .05 and .01 levels are identified by lower-case and capital letters, respectively, in the tables. No adjusted means are presented in those instances where differences in any one of the three categories were significant at the .01 level. This will preclude the possibility of attaching importance to the results of invalid comparisons. Adjusted means that were significantly different at the .05 and .01 levels are marked by d's and D's, respectively.

TABLE 12

Adjusted Performances of Samples of Examinees on the Biology Achievement Tests: Mean Scores to be Expected of Examinees Who Performed Equivalently on the Subsets of the Achievement Tests Rated Appropriate for Both Courses and on SAT-V and M

Test Date	Courses Under Comparison	Number of Examinees	Expected Mean Scores on Achievement Tests and Differences Between Means After Adjustment for Performances on: Subset of Ach. Test Means	Subset of Ach. Test Diff.	SAT-V and M Means	SAT-V and M Diff.	Both Ach. Subset and SAT-V and M Means	Both Ach. Subset and SAT-V and M Diff.
May 1962	BSCS-Blue & Reg. Biol.	30 154	511 511	0	504 512	- 8	510 511	- 1
May 1962	BSCS-Green & Reg. Biol.	80 154	488 502	-14	491 501	-10	496 498	- 2
May 1962	BSCS-Yellow & Reg. Biol.	71 154	514 519	- 5	529 512	17	518 517	1 c
May 1963	BSCS-Blue & Reg. Biol.	112 131	528 538	-10	535 532	3 b	527 539	-12
May 1963	BSCS-Green & Reg. Biol.	36 131		A		A		A
May 1963	BSCS-Yellow & Reg. Biol.	53 131	535 534	1	574 519	55 D	547 529	18 d
Jan. 1966	BSCS-Blue & Reg. Biol.	438 461	546 538	8 d	549 533	16 bD		B
Jan. 1966	BSCS-Green & Reg. Biol.	357 461	515 516	- 1	517 515	2	519 515	4 b
Jan. 1966	BSCS-Yellow & Reg. Biol.	411 461	534 526	8 cd	533 527	6	533 527	6 d
May 1966	BSCS-Blue & Reg. Biol.	463 499	561 561	0	563 558	5	559 562	- 3
May 1966	BSCS-Green & Reg. Biol.	403 499	535 536	- 1 a	531 539	- 8 a	536 535	1 a
May 1966	BSCS-Yellow & Reg. Biol.	497 499	547 552	- 5 a	551 547	4 a	548 551	- 3 ab

(Table continued on next page.)

TABLE 12 (con't)

Student-Response Samples for January and May 1967 Tests

Test Date	Courses Under Comparison	Number of Examinees	Expected Mean Scores on Achievement Tests and Differences Between Means After Adjustment for Performances on: Subset of Ach. Test Means	Subset of Ach. Test Diff.	SAT-V and M Means	SAT-V and M Diff.	Both Ach. Subset and SAT-V and M Means	Both Ach. Subset and SAT-V and M Diff.
Jan. 1967	BSCS-Blue & Reg. Biol.	500 498	518 531	-13 D	531 518	13 D	520 529	- 9 D
Jan. 1967	BSCS-Green & Reg. Biol.	500 498	501 508	- 7 d	508 501	7	504 505	- 1
Jan. 1967	BSCS-Yellow & Reg. Biol.	496 498	515 523	- 8 D	525 512	13 D	517 521	- 4
May 1967	BSCS-Blue & Reg. Biol.	500 499	534 547	-13 D	543 539	4	534 547	-13 D
May 1967	BSCS-Green & Reg. Biol.	500 499	520 524	- 4	522 522	0	521 523	- 2
May 1967	BSCS-Yellow & Reg. Biol.	500 499	533 541	- 8 D	543 531	12 D	534 540	- 6 D
School-Verified Samples for January and May 1967 Tests								
Jan. 1967	BSCS-Blue & Reg. Biol.	378 271	525 536	-11 D	524 523	11 d	525 536	-11 D
Jan. 1967	BSCS-Green & Reg. Biol.	342 271	513 514	- 1	516 511	5	513 514	- 1
Jan. 1967	BSCS-Yellow & Reg. Biol.	370 271	517 525	- 8 d	526 513	13 d	518 523	- 5
May 1967	BSCS-Blue & Reg. Biol.	420 314	540 556	-16 D	548 545	3	541 555	-14 D
May 1967	BSCS-Green & Reg. Biol.	375 314	534 534	0	535 534	1	534 535	- 1
May 1967	BSCS-Yellow & Reg. Biol.	432 314	538 545	- 7 D	545 535	10	538 544	- 6 D

[a] Differences between regular-course sample and population observed means are significantly different from differences between special-course sample and population observed means at .05 level.

[A] Differences between regular-course sample and population observed means are significantly different from differences between special-course sample and population observed means at .01 level.

[b] Regression-line slopes are significantly different at .05 level

[B] Regression-line slopes are significantly different at .01 level

[c] Standard errors of estimate are significantly different at .05 level

[C] Standard errors of estimate are significantly different at .01 level

[d] Adjusted means are significantly different at .05 level

[D] Adjusted means are significantly different at .01 level

TABLE 13

Adjusted Performances of Samples of Examinees on the Chemistry Achievement Tests: Mean Scores to be Expected of Examinees Who Performed Equivalently on the Subsets of the Achievement Tests Rated Appropriate for Both Courses and on SAT-V and M

Test Date	Courses Under Comparison	Number of Examinees	Expected Mean Scores on Achievement Tests and Differences Between Means After Adjustment for Performances on: Subset of Ach. Test Means	Subset of Ach. Test Diff.	SAT-V and M Means	SAT-V and M Diff.	Both Ach. Subset and SAT-V and M Means	Both Ach. Subset and SAT-V and M Diff.
Dec. 1961	CBA & Reg. Chem.	64 276	551 558	- 7 D				
Dec. 1961	CHEMS & Reg. Chem.	32 276	552 561	- 9 d				
Mar. 1962	CBA & Reg. Chem.	107 426	519 527	- 8 D				
Mar. 1962	CHEMS & Reg. Chem.	52 426	517 529	-12 D				
May 1962	CBA & Reg. Chem.	310 350		C				
May 1962	CHEMS & Reg. Chem.	405 350		B				
May 1963	CBA & Reg. Chem.	370 370	552 585	-33 cD				
May 1963	CHEMS & Reg. Chem.	370 370	544 585	-41 cD				
May 1964	CBA & Reg. Chem.	365 225	533 565	-32 bD				
May 1964	CHEMS & Reg. Chem.	369 225	538 570	-32 D				
Jan. 1965	CBA & Reg. Chem.	358 358		A		A		A
Jan. 1965	CHEMS & Reg. Chem.	337 358		A		A		A
May 1965	CBA & Reg. Chem.	364 370	561 571	-10 D				
May 1965	CHEMS & Reg. Chem.	370 370	568 571	- 3				
Jan. 1966	CBA & Reg. Chem.	500 500	551 560	- 9 bD				
Jan. 1966	CHEMS & Reg. Chem.	500 500	566 577	-11 D				

(Table continued on next page.)

TABLE 13 (con't)

Test Date	Courses Under Comparison	Number of Examinees	Expected Mean Scores on Achievement Tests and Differences Between Means After Adjustment for Performances on: Subset of Ach. Test Means	Subset of Ach. Test Diff.	SAT-V and M Means	SAT-V and M Diff.	Both Ach. Subset and SAT-V and M Means	Both Ach. Subset and SAT-V and M Diff.
May 1966	CBA & Reg. Chem.	496 497	550 558	- 8 d		BC		B
May 1966	CHEMS & Reg. Chem.	493 497	563 562	1		B	563 561	2 b
	Student-Response Samples for January and May 1967 Tests							
Jan. 1967	CBA & Reg. Chem.	439 500	578 574	4 d	575 577	- 2	578 574	4 d
Jan. 1967	CHEMS & Reg. Chem.	500 500		B	594 584	10 d	596 582	14 D
May 1967	CBA & Reg. Chem.	500 500	544 553	- 9 D	541 555	-14 D	544 552	- 8 D
May 1967	CHEMS & Reg. Chem.	500 500	573 567	6 d	575 565	10 d	573 567	6 D
	School-Verified Samples for January and May 1967 Tests							
Jan. 1967	CBA & Reg. Chem.	260 404	592 582	10 D	585 586	- 1	592 582	10 D
Jan. 1967	CHEMS & Reg. Chem.	371 404		B	603 590	13 d		C
May 1967	CBA & Reg. Chem.	307 415	559 565	- 6	555 568	-13 d	559 565	- 6 d
May 1967	CHEMS & Reg. Chem.	403 415	578 572	6 d	580 571	9	578 572	6 D

[a] Differences between regular-course sample and population observed means are significantly different from differences between special-course sample and population observed means at the .05 level

[A] Differences between regular-course sample and population observed means are significantly different from differences between special-course sample and population observed means at the .01 level

[b] Regression-line slopes are significantly different at .05 level

[B] Regression-line slopes are significantly different at .01 level

[c] Standard errors of estimate are significantly different at .05 level

[C] Standard errors of estimate are significantly different at .01 level

[d] Adjusted means are significantly different at .05 level

[D] Adjusted means are significantly different at .01 level

TABLE 14

Adjusted Performances of Samples of Examinees on the Physics Achievement Tests: Mean Scores to be Expected of Examinees Who Performed Equivalently on the Subsets of the Achievement Tests Rated Appropriate for Both Courses and on SAT-V and M

Test Date	Courses Under Comparison	Number of Examinees	Expected Mean Scores on Achievement Tests and Differences Between Means After Adjustment for Performances on: Subset of Ach. Test Means	Subset of Ach. Test Diff.	SAT-V and M Means	SAT-V and M Diff.	Both Ach. Subset and SAT-V and M Means	Both Ach. Subset and SAT-V and M Diff.
May '63 and Jan. '64 Tests Using Questions for Subset Chosen with Emphasis on Selecting Questions Whose Appropriateness Ratings had High Means and Low Variances								
May 1963	PSSC & Reg. Phys.	369 359	572 584	-12 cD	582 574	8 b	573 583	-10 D
Jan. 1964	PSSC & Reg. Phys.	356 363	558 556	2	551 564	-13 cd	555 559	- 4
May '63 and Jan. '64 Tests Using Questions for Subset Chosen with Emphasis on Selecting Questions with Equal Mean Appropriateness Ratings for PSSC and Regular Courses								
May 1963	PSSC & Reg. Phys.	369 359	580 576	4	582 574	8	578 578	0
Jan. 1964	PSSC & Reg. Phys.	355 363	551 563	-12 D	551 564	-13 cD	550 565	-15 bD
May 1964	PSSC & Reg. Phys.	330 337	573 567	6	570 570	0	571 569	2
Jan. 1965	PSSC & Reg. Phys.	313 332	575 577	- 2	569 582	-13 d	572 579	- 7 d
Jan. 1966	PSSC & Reg. Phys.	476 457	589 583	6 d	585 588	- 3	587 585	2
May 1966	PSSC & Reg. Phys.	514 538	623 628	- 5	620 630	-10 d	621 630	- 9 cD
Student-Response Samples for January and May 1967 Tests								
Jan. 1967	PSSC & Reg. Phys.	500 498	573 580	- 7 d	B		573 580	- 7 D
May 1967	PSSC & Reg. Phys.	500 500	588 601	-13 D	598 591	7	589 600	-11 D
School-Verified Samples for January and May 1967 Tests								
Jan. 1967	PSSC & Reg. Phys.	404 392	579 585	- 6	583 581	2	579 585	- 6 d
May 1967	PSSC & Reg. Phys.	392 392	595 608	-13 D	608 594	14 D	596 607	-11 D

[a] Differences between regular-course sample and population observed means are significantly different from differences between special-course sample and population observed means at the .05 level
[A] Differences between regular-course sample and population observed means are significantly different from differences between special-course sample and population observed means at the .01 level
[b] Regression-line slopes are significantly different at .05 level
[B] Regression-line slopes are significantly different at .01 level
[c] Standard errors of estimate are significantly different at .05 level
[C] Standard errors of estimate are significantly different at .01 level
[d] Adjusted means are significantly different at .05 level
[D] Adjusted means are significantly different at .01 level

For the May 1963 Biology Test the data are for school-verified samples; for the January and May 1967 tests in biology, chemistry, and physics data for both student-response and school-verified samples are presented. For the other tests the data are for student-response samples.

The adjustments reported in Tables 12, 13, and 14 are reasonable and proper only if fairly strong positive correlations exist between scores on the achievement tests and scores on the concomitant variables. The correlations between achievement test scores and scores on the appropriate-for-both questions cluster a little below .90. These correlations are spuriously high since the appropriate-for-both questions were in the achievement tests. The correlations between achievement scores and SAT scores tend to be around .65. The correlations were judged to be sufficiently strong to make the adjustment of achievement test scores for performance on the concomitant measures reasonable and proper.

Tables 12, 13, and 14 contain the adjusted mean scores of students of regular and special courses on six Biology Tests, nine Chemistry Tests, and eight Physics Tests. The differences between the adjusted mean scores of students from regular and special courses are categorized in Table 15 by course comparisons and by direction and magnitude. The data in Table 15 are based on student-response samples; the data for the school-verified samples in January and May, 1967 are not included in the categorization displayed in Table 15. The comparisons in physics on the May 1963 and January 1964 tests using the appropriate-for-both questions chosen with emphasis on selecting questions whose ratings had high means and low variances are also excluded. The comparisons for which no adjusted means are shown in Tables 12, 13, and 14 are not shown in Table 15.

The data in Table 15 make it clear that there is little bias in the science achievement tests. When the observed achievement means are adjusted for performance on the appropriate-for-both questions, then for none of 17 comparisons in biology, for four of 17 in chemistry, and for none of eight comparisons in physics do the differences between adjusted means of regular and special students exceed 15 scale score points. If means are adjusted for performance on SAT only, then for three of 17 comparisons

TABLE 15

Differences between Adjusted Mean Scores on Science Achievement Tests for Students of Regular and Special Courses Categorized by Course Comparisons and by Direction and Magnitude*

Course Comparisons	After Adjustment for Performance on Appropriate-for-both Questions Only							After Adjustment for Performance on SAT Only							After Adjustment for Performance on Appropriate-for-both Questions and SAT							
	Number of Regular Course Adjusted Means Higher By:			Number of Adjusted Means Equal	Number of Special Course Adjusted Means Higher By:			Number of Regular Course Adjusted Means Higher By:			Number of Adjusted Means Equal	Number of Special Course Adjusted Means Higher By:			Number of Regular Course Adjusted Means Higher By:			Number of Adjusted Means Equal	Number of Special Course Adjusted Means Higher By:		
	More than 15 pts.	5 to 15 pts.	Less than 5 pts.		Less than 5 pts.	5 to 15 pts.	More than 15 pts.	More than 15 pts.	5 to 15 pts.	Less than 5 pts.		Less than 5 pts.	5 to 15 pts.	More than 15 pts.	More than 15 pts.	5 to 15 pts.	Less than 5 pts.		Less than 5 pts.	5 to 15 pts.	More than 15 pts.
BSCS-Blue and Reg. Biology		XXX		XX		X			X			XX	XX	X		XXX	XX				
BSCS-Green and Reg. Biology		XX	XXX						XX		X	X	X				XXX		XX		
BSCS-Yellow and Reg. Biology		XXXX			X	X						X	XXX	XX		X	XX		X	X	X
CBA and Reg. Chemistry	XX	XXXXXX			X				X	X						X			X		
CHEMS and Reg. Chemistry	XX	XXX	X		X	X							XX						X	XX	
PSSC and Reg. Physics		XXXX	X		X	XX			XXX	X	X		XX			XXXXX		X	XX		

* Each X represents a difference between adjusted mean scores on a test form.

in biology, for none of four in chemistry, and for none of seven in physics do the differences between the adjusted means of regular and special students exceed 15 scale score points. If means are adjusted for performance on the appropriate-for-both questions and on SAT, then for only one of 16 comparisons in biology, for none of five in chemistry, and for none of eight in physics do the differences between the adjusted means of regular and special students exceed 15 scale score points.

The comparisons shown in Table 15 and discussed in the preceding paragraph are for student-response samples. For the January and May 1967 tests in biology, chemistry, and physics comparisons were made for both student-response samples and school-verified samples. It is assumed that comparisons based on school-verified samples are valid because these samples contain few misplaced students. If the comparisons based on student-response samples yield results closely similar to those based on school-verified samples in January and May 1967, then it is reasonable to assume that all of the earlier comparisons based on student-response samples only are as valid as the comparisons based on school-verified samples. Scrutiny of the results from the two types of samples for 1967 in Tables 12, 13, and 14 show that the results are closely similar. From the three tables one can read out for the January and May 1967 tests a total of 33 differences between adjusted means based on student-response samples and contrast these with a total of 33 differences between adjusted means based on the corresponding school-verified samples. In eight cases the differences based on the two types of samples are identical, in 21 cases the differences differ by 4 points or less, and in only four instances do the differences differ by more than 4 scale score points. It seems entirely reasonable to conclude that the comparisons based on student-response samples are as valid as comparisons based on school-verified samples.

In chemistry, considering only student-response samples, there were 17 valid comparisons (nine between CBA and regular and eight between CHEMS and regular) in which adjustments were made for performance on appropriate-for-both questions, but only five valid comparisons (two between CBA and regular and three between CHEMS and regular) in which adjustments were made for performance on appropriate-for-both questions and SAT. This procedure resulted from the thought that the

appropriate-for-both adjustment was the more appropriate, critical, and meaningful one to make. The data in Table 13 show that the adjustments using appropriate-for-both questions only yield results nearly identical to those obtained when adjustments for all concomitant variables are made. In no instance do the differences between adjusted means resulting from adjusting for only the one concomitant variable differ by more than 1 point from the differences between means resulting from adjusting for all concomitant variables.

In biology and physics as well the adjustments for all concomitant variables yield results similar to the results of adjusting only for performance on the appropriate-for-both questions. However, the similarity is not quite as striking as that observed for chemistry. If one looks only at the January and May 1967 results for school-verified samples, it is clearly apparent that the adjustments involving all concomitant variables are much more closely similar to the adjustments involving the appropriate-for-both questions only than to those using SAT only. This is to be expected since correlations between the achievement test scores and appropriate-for-both question scores are higher than correlations between achievement test scores and SAT scores.

A final topic for discussion is the differing adjustments that resulted for the May 1963 and January 1964 Physics Tests from using two different sets of appropriate-for-both questions. When the Physics Test means were adjusted for performance on a set of appropriate-for-both questions chosen with emphasis on picking questions that had PSSC and regular teacher ratings with high means and low variances, the regular adjusted mean exceeded the PSSC mean on the May 1963 test by 12 points, whereas the PSSC mean exceeded the regular mean by 2 points on the January 1964 test. When the emphasis was on picking questions with equal PSSC and regular ratings, the PSSC mean exceeded the regular mean by 4 points on the May 1963 test and the regular mean exceeded the PSSC mean by 12 points on the January 1964 test. These results indicate the sensitivity of the statistical adjusting procedure to the questions selected to serve as the course-free measure of science achievement. As discussed in Part C on teacher ratings, the procedure ultimately adopted took both considerations into account. Questions with high and equal regular and special ratings were selected for the measures of course-free science achievement.

VI. CONCLUSIONS

The evidence clearly supports the general conclusion that the College Board Science Achievement Tests are equally appropriate for students of regular and special courses in biology, chemistry, and physics.

The appropriateness of six Biology Tests for BSCS-Blue and Regular Biology was studied. On the first two tests there were indications from teacher ratings of bias in favor of Regular Biology, but these were not borne out by analyses of student test performance. On five of the six tests studied, the BSCS-Blue means were higher than the Regular Biology means. After adjusting for performance on the questions rated appropriate-for-both BSCS-Blue and regular, then none out of the six tests was found to be biased. An adjustment using all concomitant variables was excluded for one test because the regression-line slopes were significantly different at the .01 level. The adjustments on the remaining five tests using all concomitant variables revealed no bias.

The appropriateness of six Biology Tests for BSCS-Green and Regular Biology was studied. On the first test, there was an indication from teacher ratings of bias in favor of Regular Biology, but this was not borne out by analysis of student test performance. The adjusted student performance on one of the six tests was excluded because the difference between the BSCS-Green population and sample means was significantly different at the .01 level from the difference between the Regular Biology population and sample means. On the remaining five Biology Tests the Regular Biology means were higher than BSCS-Green means in every case. Regular Biology means were higher than BSCS-Green means on all concomitant measures, also, in every case. After adjustment for performance on concomitant measures, none of the tests was found to be biased.

The appropriateness of six Biology Tests for BSCS-Yellow and Regular Biology was studied. On the first two tests, there were indications from teacher ratings of bias in favor of Regular Biology, but these were not borne out by analysis of student test performance. In fact, on the second test, the Regular Biology students scored considerably higher on SAT-V and M, but the BSCS-Yellow students scored higher on the

Biology Test. The analysis of covariance adjusting for all concomitant measures showed that this test was biased in favor of BSCS-Yellow. On all six Biology Tests, BSCS-Yellow means exceeded Regular Biology means. Adjustment only for performance on questions rated appropriate for both courses revealed no bias in any of the six tests. Teacher ratings indicated the fourth test was biased in favor of BSCS-Yellow.

The appropriateness of 12 Chemistry Tests for CBA and Regular Chemistry was studied. The teacher ratings indicated that the fourth, fifth, and sixth tests were biased in favor of Regular Chemistry. On the fifth test no data on student performance are available, but on the fourth and sixth tests analysis of student performance clearly supported the teacher judgments. On these two Chemistry Tests, the regular students scored considerably higher, even though the CBA students scored higher or nearly as high on the questions rated appropriate for both courses. The performances of CBA and Regular Chemistry students on 11 Chemistry Tests were analyzed. One of these tests was excluded because the samples were not representative of the populations. On the ten remaining tests the means of Regular Chemistry students surpassed the means of CBA students in every case. Valid adjustments for performance on the questions rated appropriate for CBA and regular revealed two biased tests (the fourth and sixth); both favored Regular Chemistry. Means on four Chemistry Tests were adjusted for performance on all concomitant measures. One of those tests was the one already excluded because of nonrepresentative samples. Another was excluded because of significantly different regression-line slopes. For neither of the two remaining tests did adjustment for all concomitant variables reveal any bias. For the last five tests studied, there were no indications of bias from any source.

The appropriateness of 12 Chemistry Tests for CHEMS and Regular Chemistry was studied. Evidence from teacher ratings indicated that the fourth, fifth, and sixth tests were biased in favor of Regular Chemistry, whereas the tenth test was biased in favor of CHEMS. Actual performance on the fourth and sixth tests supported the teacher judgments; the Regular Chemistry means on the Chemistry Tests were considerably higher than CHEMS means even though the regular and CHEMS students performed at nearly the same level on the questions rated appropriate for both courses.

Student performance on the fifth test was not analyzed. For the 11 tests on which student performances were studied, one test was excluded because of nonrepresentative samples. On eight of the ten remaining tests the means of Regular Chemistry students were higher than the means of CHEMS students. Adjustments for performances on appropriate-for-both questions were ruled out on two tests because of significantly different regression-line slopes. On the eight remaining tests adjustments for performances on appropriate-for-both questions revealed bias in favor of Regular Chemistry in the case of two tests; these were the fourth and sixth tests. Means on four Chemistry Tests were adjusted for performance on all concomitant measures. One of these tests was the one already excluded because of unrepresentative samples. No bias was found in the three remaining tests after adjusting for all concomitant variables.

The last Chemistry Test showing bias from adjusted performance data in favor of Regular Chemistry was introduced in May 1964. Five Chemistry Tests introduced since then have shown no bias except for an indication from actual performance of bias in favor of regular over CHEMS for the January 1966 test and an indication from teacher ratings of bias in favor of CHEMS over regular in the May 1966 test. Adjustments using only the appropriate-for-both questions and adjustments using all concomitant variables yield nearly identical results in chemistry. Hence, the small number of Chemistry Tests for which adjustments using all concomitant variables were made does not cast much, if any, doubt on the general conclusion that the most recent Chemistry Tests are unbiased.

The appropriateness of eight Physics Tests for PSSC and Regular Physics was studied. Evidence from teacher ratings indicated bias in the fourth and sixth tests in favor of Regular Physics. One of these indications was supported by actual but not adjusted performance data. The performances of PSSC and Regular Physics students on eight Physics Tests were studied. On seven of the eight tests the PSSC means exceeded the Regular Physics means. After adjusting for performance on the appropriate-for-both questions and again after adjusting for performance on all concomitant variables, no examples of bias were found.

In summary, the indications of bias in the Biology and Physics Tests were few in number. The few indications that there were stemmed almost exclusively from

teacher ratings of appropriateness, and these teacher ratings were not borne out by student performance. Both teacher ratings and analyses of student performance indicated that Chemistry Tests introduced in 1963 and 1964 were biased in favor of Regular Chemistry over CBA and CHEMS. On five Chemistry Tests introduced since then, however, there were no indications of bias in favor of regular over CBA; there was one indication from actual performance of bias in favor of regular over CHEMS; but one indication, also, from teacher ratings of bias in favor of CHEMS over regular. There were no indications of bias from adjusted performance data on the last five Chemistry Tests studied.

LIST OF FIGURES

Figure	Title	Page
1	Examples of Various Relationships Between Course X Sample and Population Means and Course Y Sample and Population Means	13
2	Numbers of Biology Examinees in Different Biology Courses	16
3	Numbers of Chemistry Examinees in Different Chemistry Courses	17
4	Numbers of Physics Examinees in Different Physics Courses	18
5	Biology — Mean Scores on the Complete Biology Achievement Tests of Populations and Samples of Examinees from Different Biology Courses	38
6	Chemistry — Mean Scores on the Complete Chemistry Achievement Tests of Populations and Samples of Examinees from Different Chemistry Courses	39
7	Physics — Mean Scores on the Complete Physics Achievement Tests of Populations and Samples of Examinees from Different Physics Courses	41
8	Biology: BSCS-Blue and Regular Sample Means on the Complete Biology Achievement Tests, on Subsets of the Biology Achievement Tests Rated Appropriate for Both Courses, and on SAT-V and M	43
9	Biology: BSCS-Green and Regular Sample Means on the Complete Biology Achievement Tests, on Subsets of the Biology Achievement Tests Rated Appropriate for Both Courses, and on SAT-V and M	44
10	Biology: BSCS-Yellow and Regular Sample Means on the Complete Biology Achievement Tests, on Subsets of the Biology Achievement Tests Rated Appropriate for Both Courses, and on SAT-V and M	46
11	Chemistry: CBA and Regular Sample Means on the Complete Chemistry Achievement Tests, on Subsets of the Chemistry Achievement Tests Rated Appropriate for Both Courses, and on SAT-V and M	47
12	Chemistry: CHEMS and Regular Sample Means on the Complete Chemistry Achievement Tests, on Subsets of the Chemistry Achievement Tests Rated Appropriate for Both Courses, and on SAT-V and M	48
13	Physics: PSSC and Regular Sample Means on the Complete Physics Achievement Tests, on Subsets of the Physics Achievement Tests Rated Appropriate for Both Courses, and on SAT-V and M	49

LIST OF TABLES

Table	Title	Page
1	Textbooks Used in Regular Courses by Samples of Mid-School-Year (1965-1966) Senior Examinees and End-of-School-Year (1965-1966) Junior Examinees in Biology, Chemistry, Physics . . .	20
2	Accuracy of Examinee Responses Regarding BSCS Biology Courses Studied — May 1963 Biology Test	21
3	Accuracy of Examinee Responses Regarding Physics Courses Studied — May 1963 and January 1964 Physics Tests	22
4	Accuracy of Examinee Responses Regarding Science Courses Studied — January and May 1967 Biology, Chemistry, and Physics Tests	23
5	Biology — Teacher Ratings of Appropriateness	26
6	Chemistry — Teacher Ratings of Appropriateness	27
7	Physics — Teacher Ratings of Appropriateness	30
8	Actual Performance of Samples of Biology Examinees on the Biology Achievement Tests, on Subsets of Questions in the Biology Achievement Tests Rated Appropriate for Both Courses Under Comparison, and on the Scholastic Aptitude Test	32
9	Actual Performance of Samples of Chemistry Examinees on the Chemistry Achievement Tests, on Subsets of Questions in the Chemistry Achievement Tests Rated Appropriate for Both Courses Under Comparison, and on the Scholastic Aptitude Test	34
10	Actual Performance of Samples of Physics Examinees on the Physics Achievement Tests, on Subsets of Questions in the Physics Achievement Tests Rated Appropriate for Both Courses Under Comparison, and on the Scholastic Aptitude Test	36
11	Actual Performance of Student-Response and School-Verified Samples of Biology, Chemistry, and Physics Examinees from January and May 1967 on the Biology, Chemistry, and Physics Achievement Tests, respectively; on Subsets of the Biology, Chemistry, and Physics Tests Rated Appropriate for Courses under Comparison, respectively; and on the Scholastic Aptitude Test	51
12	Adjusted Performances of Samples of Examinees on the Biology Achievement Tests — Mean Scores to be Expected of Examinees who Performed Equivalently on the Subsets of the Achievement Tests Rated Appropriate for both Courses and on SAT-V and M. . .	53

LIST OF TABLES (Cont'd)

Table	Title	Page
13	Adjusted Performances of Samples of Examinees on the Chemistry Achievement Tests — Mean Scores to be Expected of Examinees who Performed Equivalently on the Subsets of the Achievement Tests Rated Appropriate for both Courses and on SAT-V and M	55
14	Adjusted Performances of Samples of Examinees on the Physics Achievement Tests — Mean Scores to be Expected of Examinees who Performed Equivalently on the Subsets of the Achievement Tests Rated Appropriate for Both Courses and on SAT-V and M . . .	57
15	Differences between Adjusted Mean Scores on Science Achievement Tests for Students of Regular and Special Courses Categorized by Course Comparisons and by Direction and Magnitude . .	59

APPENDIX I

Form PAC1

SCIENCE ACHIEVEMENT TESTS

Question Rating Form

I have rated the questions in terms of their appropriateness for:

Modern Biology _____
BSCS (Blue Version) _____
BSCS (Green Version) _____
BSCS (Yellow Version) _____

Modern Chemistry _____
CBA _____
CHEMS _____

Modern Physics _____
PSCS _____

No.	Answer	Appropriate And Emphasized	Appropriate But Not Emphasized	In-appropriate	No.	Answer	Appropriate And Emphasized	Appropriate But Not Emphasized	In-appropriate
1.					22.				
2.					23.				
3.					24.				
4.					25.				
5.					26.				
6.					27.				
7.					28.				
8.					29.				
9.					30.				
10.					31.				
11.					32.				
12.					33.				
13.					34.				
14.					35.				
15.					36.				
16.					37.				
17.					38.				
18.					39.				
19.					40.				
20.					41.				
21.					42.				

APPENDIX II

DIRECTIONS FOR RATING TEST QUESTIONS

Check the course for which you are rating the appropriateness of the questions.

For each question, enter the answer (A, B, C, D, or E) in the column headed Answer and then place a check in one of the three columns that follow.

Check the column headed Appropriate and Emphasized if the question is based on material that is emphasized in the course for which you are rating the appropriateness of the questions.

Check the column headed Appropriate But Not Emphasized if the question is one which you think students who have completed the course should be able to answer even though the question is based on material not emphasized in the course.

Check the column headed Inappropriate if you think that students who have completed the course should not be expected to know the answer to the question.

APPENDIX III

Total Numbers of Examinees Taking Each Science Achievement Test per School Year

Test	Year					
	1962-63	1963-64	1964-65	1965-66	1966-67	1967-68
Biology	32,888	41,270	48,894	50,506	52,613	60,776
Chemistry	53,156	61,110	65,729	66,997	65,628	67,816
Physics	28,466[a]	29,192	30,076	28,528	28,435	29,414

[a]Includes 27,152 examinees who took the Physics Achievement Test and 1,314 who took the PSSC Physics Achievement Test

APPENDIX IV

Numbers and Percentages of Examinees in Different Courses

Course	May 1962	Jan. 1963	May 1963	Jan. 1964	May 1964	Jan. 1965
			Biology			
BSCS Blue	58 (1%)	51 (1%)	256 (2%)	162 (2%)	481 (3%)	399 (3%)
BSCS Green	136 (1)	34 -	98 (1)	241 (3)	532 (3)	340 (3)
BSCS Yellow	104 (1)	52 (1)	174 (1)	106 (1)	642 (4)	235 (2)
BSCS (not sure of Version)	--	69 (1)	89 (1)	297 (3)	486 (3)	446 (4)
Not BSCS	10,770 (97)	7,388 (97)	7,007 (51)	5,060 (55)	8,033 (50)	6,294 (55)
Not Sure[a]			6,160 (45)	2,010 (22)	3,399 (21)	2,671 (23)
No Data[b]				1,398 (15)	2,393 (15)	1,039 (9)
Total	11,068	7,594	13,784	9,274	15,966	11,424
			Chemistry			
CBA			560 (2%)	367 (3%)	878 (3%)	517 (4%)
CHEMS			1,472 (5)	772 (7)	3,495 (13)	1,731 (13)
Not Sure[a]			6,603 (24)	3,324 (30)	8,419 (30)	3,943 (30)
Not CBA or CHEMS			11,611 (43)	5,684 (50)	12,414 (45)	6,342 (48)
No Data[b]			7,045 (26)	1,069 (10)	2,443 (9)	613 (5)
Total			27,291	11,216	27,649	13,146
			Physics			
PSSC	(22%)	1,473 (21%)	2,121 (22%)	2,036 (26%)	2,608 (32%)	2,640 (31%)
Not Sure[a]		1,086 (16)	1,262 (13)	1,181 (15)	1,165 (14)	1,583 (18)
Not PSSC	(47)	3,395 (49)	4,456 (47)	3,825 (49)	3,694 (45)	4,006 (47)
No Data[b]		1,023 (15)	1,595 (17)	727 (9)	684 (8)	351 (4)
Total	8,672	6,977 (101)[c]	9,434	7,769	8,151	8,580

APPENDIX IV (con't)

Course	Test Dates				
	May 1965	Jan. 1966	May 1966	Jan. 1967	May 1967
Biology					
BSCS Blue	1,247 (7%)	588 (5%)	1,262 (8%)	798 (7%)	1,588 (10%)
BSCS Green	876 (5)	436 (4)	1,022 (6)	669 (6)	1,213 (7)
BSCS Yellow	1,571 (9)	514 (5)	2,375 (15)	1,091 (9)	2,848 (17)
BSCS (not sure of Version)	621 (4)	483 (4)	694 (4)	642 (6)	830 (5)
Not BSCS	7,452 (45)	5,480 (49)	6,203 (38)	4,855 (42)	5,207 (32)
Not Sure[a]	3,347 (20)	2,504 (23)	2,902 (18)	2,536 (22)	2,845 (17)
No Data[b]	1,634 (10)	1,122 (10)	1,665 (10)	931 (8)	1,980 (12)
Total	16,739	11,097	16,123	11,522	16,511
Chemistry					
CBA	1,093 (4%)	527 (4%)	1,121 (4%)	510 (4%)	964 (4%)
CHEMS	5,007 (19)	2,295 (18)	6,537 (25)	2,833 (24)	8,121 (32)
Not Sure[a]	7,299 (28)	3,627 (29)	6,566 (25)	3,522 (29)	6,431 (25)
Not CBA or CHEMS	11,585 (44)	5,057 (41)	9,894 (38)	4,371 (36)	8,215 (32)
No Data[b]	1,324 (5)	900 (7)	1,877 (7)	767 (6)	1,917 (7)
Total	26,308	12,406	25,995	12,003	25,648
Physics					
PSSC	2,786 (36%)	2,636 (32%)	2,504 (38%)	3,067 (37%)	2,629 (38%)
Not Sure[a]	1,288 (17)	1,421 (17)	990 (15)	1,489 (18)	1,041 (15)
Not PSSC	3,335 (43)	3,499 (43)	2,685 (40)	3,239 (39)	2,651 (38)
No Data[b]	339 (4)	629 (7)	483 (7)	600 (7)	565 (8)
Total	7,748	8,185	6,662	8,395	6,886

[a]Examinees in the "Not Sure" category are those who indicated that they were not sure about the course they took.

[b]Examinees in the "No Data" category included a few who marked the spaces to be left blank and those who failed to respond altogether.

[c]Percentages do not total 100% in all cases because of rounding to the nearest whole number.

APPENDIX V

Mean Test Scores of Students in Different Course Categories

Course	Test Dates										
	May '62	Jan. '63	May '63	Jan. '64	May '64	Jan. '65	May '65	Jan. '66	May '66	Jan. '67	May '67
					Biology						
BSCS Blue	511	538	541	549	536	562	540	561	575	535	550
BSCS Green	476	469	505	490	488	486	469	506	526	498	524
BSCS Yellow	512	537	555	519	562	539	540	535	557	524	562
BSCS (not sure of Version)	---[a]	499	465	466	470	465	480	470	490	455	494
Not BSCS	---[a]	---[b]	538	528	542	525	536	529	539	511	532
Not Sure	---[a]	---[b]	---[c]	490	508	484	504	489	511	474	510
No Data	---[a]	---[b]	---[c]	474	477	462	469	460	479	462	480
Grand Mean	515	513	519	508	521	507	520	511	530	496	527
					Chemistry						
CBA			557	534	521	544	545	536	547	568	539
CHEMS			545	538	537	576	571	566	561	598	576
Not Sure			545	512	532	515	542	513	525	536	528
Not CBA or CHEMS			581	562	573	540	581	571	566	580	568
No Data			532	510	508	512	529	529	506	556	515
Grand Mean			557	539	548	550	564	548	549	569	556
					Physics						
PSSC	591	567	585	562	576	579	604	595	625	588	617
Not Sure	---[d]	523	525	516	522	539	550	546	576	542	557
Not PSSC	559	558	559	558	555	579	584	582	610	572	590
No Data	---[d]	508	508	513	502	523	514	558	535	559	538
Grand Mean	554	547	551	548	552	569	582	578	605	570	591

[a] The mean of the group including students in the last four categories on the May 1962 Biology Test was 515.
[b] The mean of the group including students in the last three categories on the January 1963 Biology Test was 513.
[c] The mean of the group including students in the "Not Sure" and the "No Data" categories for the May 1963 Biology Test was 474.
[d] The mean of the group including students in the "Not Sure" and the "No Data" categories for the May 1962 Physics Test was 521.

BIBLIOGRAPHY

1. Angoff, William H. Curriculum Bias in the Achievement Tests. in Angoff, W. H., ed., The College Board Admissions Testing Program: A Technical Report on Research and Development Activities Relating to the Scholastic Aptitude Test and Achievement Tests. Princeton, N. J.: Educational Testing Service, 1971, pp. 57-58.

2. Coffman, William E. The Purposes of the Achievement Tests. in Angoff, W. H., ed., The College Board Admissions Testing Program: A Technical Report on Research and Development Activities Relating to the Scholastic Aptitude Test and Achievement Tests. Princeton, N. J.: Educational Testing Service, 1971, pp. 49-50.

3. Fornoff, Frank J. A Survey of the Teaching of Chemistry in Secondary Schools. Test Development Report, TDR-69-5, Educational Testing Service, May 1969.

4. Fornoff, Frank J. Developing the New Physics Test. College Board Review, No. 46, 1962.

5. Kastrinos, William. A Survey of the Teaching of Biology in Secondary Schools. Test Development Report, TDR-69-6, Educational Testing Service, May 1969.

6. Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education, Houghton Miflin Company, Boston, 1953.

7. Malcolm, D. J. An analysis of the performance of the candidates from BSCS and conventional biology courses on the CEEB Biology Test, Form KAC1. Unpublished report, Educational Testing Service, 1962.

8. Pitcher, Barbara. College Entrance Examination Board, Candidates and Tests, 1963-1969. College Entrance Examination Board, 1963-1969.

9. Stickell, David W. A Comparison of the Performance of BSCS and Non-BSCS Candidates on the CEEB Biology Achievement Test May 1963 Administration. Test Development Report, TDR-65-1, Educational Testing Service, March 1965.

10. Swineford, Frances. College Entrance Examination Board January 1962 through January 1969 Achievement Examinations. Test Analyses, SR-62-37 through SR-69-52, Educational Testing Service, July, 1962 through June, 1969.

11. Thompson, Raymond E. A Survey of the Teaching of Physics in Secondary Schools. Test Development Report, TDR-69-3, Educational Testing Service, January 1969.